

Advanced Sciences and Technologies for Security Applications

Adam Henschke  
Alastair Reed  
Scott Robbins  
Seumas Miller *Editors*

# Counter-Terrorism, Ethics and Technology

Emerging Challenges at the Frontiers  
of Counter-Terrorism

OPEN ACCESS

 Springer

# **Advanced Sciences and Technologies for Security Applications**

## **Series Editor**

Anthony J. Masys, Associate Professor, Director of Global Disaster Management, Humanitarian Assistance and Homeland Security, University of South Florida, Tampa, USA

## **Advisory Editors**

Gisela Bichler, California State University, San Bernardino, CA, USA

Thirimachos Bourlai, Lane Department of Computer Science and Electrical Engineering, Multispectral Imagery Lab (MILab), West Virginia University, Morgantown, WV, USA

Chris Johnson, University of Glasgow, Glasgow, UK

Panagiotis Karampelas, Hellenic Air Force Academy, Attica, Greece

Christian Leuprecht, Royal Military College of Canada, Kingston, ON, Canada

Edward C. Morse, University of California, Berkeley, CA, USA

David Skillicorn, Queen's University, Kingston, ON, Canada

Yoshiki Yamagata, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan

## Indexed by SCOPUS

The series *Advanced Sciences and Technologies for Security Applications* comprises interdisciplinary research covering the theory, foundations and domain-specific topics pertaining to security. Publications within the series are peer-reviewed monographs and edited works in the areas of:

- biological and chemical threat recognition and detection (e.g., biosensors, aerosols, forensics)
- crisis and disaster management
- terrorism
- cyber security and secure information systems (e.g., encryption, optical and photonic systems)
- traditional and non-traditional security
- energy, food and resource security
- economic security and securitization (including associated infrastructures)
- transnational crime
- human security and health security
- social, political and psychological aspects of security
- recognition and identification (e.g., optical imaging, biometrics, authentication and verification)
- smart surveillance systems
- applications of theoretical frameworks and methodologies (e.g., grounded theory, complexity, network sciences, modelling and simulation).

Together, the high-quality contributions to this series provide a cross-disciplinary overview of forefront research endeavours aiming to make the world a safer place.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors.

More information about this series at <https://link.springer.com/bookseries/5540>

Adam Henschke · Alastair Reed · Scott Robbins ·  
Seumas Miller  
Editors

# Counter-Terrorism, Ethics and Technology

Emerging Challenges at the Frontiers  
of Counter-Terrorism

 Springer

*Editors*

Adam Henschke  
Philosophy Section  
University of Twente  
Enschede, Netherlands

Alastair Reed  
Cyber Threats Research Centre  
Swansea University  
Swansea, UK

Scott Robbins  
Center for Advanced Security, Strategic  
and Innovation Studies (CASSIS)  
University of Bonn  
Bonn, Germany

Seumas Miller  
Charles Sturt University  
Canberra, Australia  
TU Delft  
Delft, Netherlands

University of Oxford  
Oxford, England



ISSN 1613-5113

ISSN 2363-9466 (electronic)

Advanced Sciences and Technologies for Security Applications

ISBN 978-3-030-90220-9

ISBN 978-3-030-90221-6 (eBook)

<https://doi.org/10.1007/978-3-030-90221-6>

© The Editor(s) (if applicable) and The Author(s) 2021. This book is an open access publication.

**Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Acknowledgments

The research was conducted under the auspices of the European Research Council's Advanced Grant program as part of the grant entitled, "Global Terrorism and Collective Moral Responsibility: Redesigning Military, Police and Intelligence Institutions in Liberal Democracies" (GTCMR. No. 670172) (Principal Investigator: Professor Seumas Miller; institutional partners, Delft University of Technology and the University of Oxford). The Australian Research Council Discovery Grant, entitled, "Intelligence And National Security: Ethics, Efficacy And Accountability" (DP180103439) (Principal Investigator Professor Seumas Miller; institutional partners Charles Sturt University and the Australian National University). The Australian Department of Defence Strategic Policy Grant, entitled "Countering Foreign Interference And Cyber War Challenges" (Principal Investigator Dr Adam Henschke; institutional partner, the Australian National University).

# Introduction

On April 2019, the terrorist group National Thawheed Jamaath carried out a lethal terrorist attack in Sri Lanka that targeted Christians on Easter Sunday. “In all, eight men and one woman belonging to local Islamist groups detonated bombs almost simultaneously in several parts of the country, killing themselves and more than 250 others” [1]. Following the attacks, the Sri Lankan Defence Minister stated that the attacks were a response to the terrorist attack in Christchurch, New Zealand, in which the gunman killed more than 50 Muslims [2]. These attacks led the New Zealand Prime Minister Jacinta Ardern and French President Emmanuel Macron to lobby social media companies around the world to do more in the fight against terrorism. Ardern stated that “[t]his isn’t about freedom of expression; this is about preventing violent extremism and terrorism online” [3]. This shows how terrorism is now truly international—a terrorist attack conducted by an Australian in New Zealand against Muslims is said to have led to a terrorist attack in Sri Lanka conducted against Christians, driving two world leaders to seek changes from technology companies. Moreover, it shows that technology is now as much a part of terrorism and counter-terrorism as it is for all other parts of modern life. To say that we need to understand and respond to these new forms of violent extremism is obvious. The ethics of *how* we respond are complex and varied.

The 2001 al Qaeda attacks on the USA caused a seismic shift in how the world viewed terrorism. Then, in 2013 Edward Snowden released a trove of data that gave the world a glimpse of the technological power being wielded in the name of counter-terrorism. Since then we have witnessed the rise of social media being used by terrorists and counter-terrorist agencies, unprecedented hype for the use of artificial intelligence and machine learning for counter-terrorism, and the practices falling under bulk data collection ever increasing. Moreover, we are also bearing witness to a range of technologies being extended in their use as part of counter-terrorist practices—from the use of facial recognition technologies, to the ways we respond to weapons of mass destruction, to the development of social credit systems as tools for population control, justified by reference to the needs of counter-terrorism. This edited volume takes stock of the recent evolution of international terrorism, the development of modern technologies, and modernisation of more long-standing

technologies are being used to counter terrorists—and how terrorist organisations are leveraging it for their own purposes.

Getting a handle on how these technologies are actually used in the context of terrorism and counter-terrorism offers a way to start to separate the hype from the reality. For example, although there is much discussion of cyber-terrorism, there is little real-world activity that falls under this heading. This does not mean that important activities are not taking place with the aid of modern technologies. Terrorists are using drones to attack government forces, using social media for propaganda and recruitment, and encryption to evade detection. Counter-intelligence agencies are using machine learning to detect suspicious behaviour, hacking computers to gain access to encrypted data, and collecting bulk data in quantities too large to describe. The Christchurch shootings were notable as the shooter not only livestreamed his attacks, using social media to broadcast the attacks as they occurred, but also paired his attacks with an online manifesto, that has subsequently been linked to a range of nationalist terrorist efforts. Again, we see how new technologies and the social behaviours associated with them are evolving in parallel with terrorism.

Moreover, we also need to consider how terrorist use of technologies and the counter-terrorism responses impact the wider society. Social media is now a fundamental part of modern life—woven into people’s personal lives, communities, and political activity. While many people might agree that social media companies ought to do more combat terrorist use of their tools, we must also confront concerns about free speech, free association, and the overreach of government. While the uses of these technologies are interesting in and of themselves, it is important to know whether or not these technologies are effective at countering terrorism. How often is a machine learning algorithm correct when it tags someone as suspicious? And how many terrorists does it miss? How does this compare to the old way of countering terrorism? Each technology and its application has its own set of difficulties when evaluating it for efficacy. This volume provides insight into either how efficacious these technologies are, or how we can go about evaluating them for efficacy. This efficacy is a key component of any ethical assessment of a counter-terrorism technology.

Ultimately, the questions here touch on deep ethical issues. What we mean here is twofold—first, when considering the adoption of a set of technologies like drones in the fight against terrorists, the use of surveillance to place individuals and groups under constant government observation, or whether encryption technologies should be used by citizens or “cracked” by counter-terrorism agencies, we are engaging with ethical content. Should drones be used at all? Is government surveillance permissible or is it a violation of individual privacy? Will the loss of encryption technology undermine the security of the internet, and should we care? Second, each of these questions requires us to engage in ethical reflection. What we mean here is that we cannot simply pass judgement on these actions by governments and individuals, deeming them right or wrong. We need to actively *reflect* upon those judgements, to look at the reasons that underpin them, to see if the actions, the judgements, and responses can be justified or not.



Finally, while some of these technologies may be effective means of counter-terrorism, we must go a step further and ensure that it is ethical to use technologies for these purposes. Much has been discussed about the ethics surrounding so-called killer robots—for counter-terrorism and warfare. Little academic discussion has been focused on other technologies. This is unfortunate as technologies like facial recognition technology, bulk data collection, and social media are actually being used today to counter terror. Each of these technologies present novel ethical issues which must be understood if we are to ensure that liberal democratic values are preserved while countering terrorism.

This book grapples with these ethical issues sitting at the frontiers of counter-terrorism, covering a range of different technologies and practices that span terrorism, counter-terrorism, and modern social practices. The threads are somewhat disparate, but weave together a story of similar challenges—how are technologies changing terrorist behaviours, driving the responses by counter-terrorism, and what are the criticisms and justifications for those behaviours and responses? The book comes in five main parts, each looking at different threads of this larger story.

The first part *Understanding Counter-Terrorism Technologies: Drones and the Ethical Risks of New Technologies* looks at how technologies shape the practice and understanding of counter-terrorism, looking at one of the most controversial sets of technologies used in efforts against terrorists: drones. Jessica Wolfendale starts by conceptualising the notion of terrorism and shows how state actions with particular technologies including drones offer significant ethical risks. Michael Robillard then looks at the relation between drone use and the narratives that develop around counter-terrorism practices. Amanda and Noel Sharkey then discuss the ways that terrorists and others can exploit particular features of drones, in service of their larger political aims.

Concepts of terrorism and technology are fundamental to any discussion of the ethics of counter-terrorism. In *Technology as Terrorism: Police Control Technologies and Drone Warfare*, Jessica Wolfendale presents an argument that technology, and the language we use to talk about technology, constrains and shapes our moral understanding of the nature, scope, and impact of terrorism, particularly in relation to state terrorism. This chapter offers conceptual discussions of the notion of terrorism, and the relation to state use of police control and drone technologies are combined with a narrative of precision and efficiency. This language masks the terroristic nature of the violence that these practices inflict and reinforces the moral exclusion of those against whom these technologies are deployed.

Michael Robillard also looks at drone technologies, but focusses attention on how the use of drone technologies in counter-terrorism operations bear upon the larger campaign to “win hearts and minds”. He argues that an underlooked aspect of the use of drones in counter-terrorism operations is proper regard for the moral significance that the non-kinetic features of narrative, imagery, and social signalling play with respect to remote targeted killing operations. A fundamental aspect of any effective counter-terrorism operation is the narrative that goes along with that, and the use of relatively new technologies like drones must be seen with regard to that narrative.

In another approach to the ethical use of drones in counter-terrorism operations, Amanda Sharkey and Noel Sharkey look at how “deception” of these autonomous weapon systems is an increasingly important element. The basic worry here is that the absence of human control of autonomous weapon systems necessitates a changed perspective on the notion of deception that has not yet made its way into military manuals. They ask how does deception fit into the ongoing technological transformation of warfare where ever more control of weapons is being ceded to computer systems?

In the second part, *The Challenges of Technologies of Terrorism and Counter-Terrorism: Weapons of Mass Destruction, the Internet of Things, and Facial Recognition Technology* offers analysis of three different technology types to show how use of particular technology types presents challenges for counter-terrorism. Jonas Feltes begins the section looking at the notion of weapons of mass destruction (WMD), to offer an argument for a new way of considering these technologies. Adam Henschke then suggests that the Internet of Things (IoT) will usher in a new era where cyber-terrorism will present risks in the physical world, requiring us to anticipate this emergent risk and to prepare for it. Scott Robbins closes this part out by looking at Facial Recognition Technologies (FRT) to offer a set of arguments why some restrictions on the use of FRT for CT are ethically justified.

One of the deepest concerns that has driven a considerable aspect of counter-terrorism policies is what happens if a terrorist group has access to and uses a weapon of WMD. Jonas Feltes drills down into these concerns by providing a critical engagement of the concept of WMDs, showing the relations between chemical, biological, radiological, or nuclear weapons technologies the general concept of WMD. He argues that a static concept that includes or excludes certain weapon types purely on the basis of their physical impact in an attack deals with problematic threshold issues and ethical challenges. He instead offers a complex understanding of the impact of particular weapons, their availability to terrorists, such that the threat that terrorist attacks with improvised unconventional weapons can be analysed and displayed more accurately. This more nuanced approach both allows for more efficacious and precise counter-terrorism practices and policy and can reduce ethically unsustainable behaviour of first responders and the press during a terrorist incident.

Adam Henschke next looks at the IoT, the cluster of technologies that span the cyber and physical realms. In this chapter, he argues that this blurring and integration of the cyber and physical realms means that cyber-terrorism will take place. The threat of terrorism is an emergent threat, arising from the combination of five related features of the IoT: it is radically insecure, its components are in the world, that the sheer numbers of IoT devices mean potential attacks can be intense, its reliance on artificial intelligence will make aspects of it inscrutable, and that the IoT is largely invisible. As the IoT grows in scope and penetration of our physical worlds and behaviours, it means that cyber-terrorism is not a question of if, but when. This has significant ethical implications as these five features of the IoT mean that we ought to be regulating these technologies.

FRT is the third set of technologies requiring an ethical analysis. In this chapter, Scott Robbins explores the ways that FRT is used as part of counter-terrorism practices. Working from the recognition that while FRT might be justifiable, five conditions must be met for it to be ethically permissible. First, the state must create institutional constraints that only allow FRTs to be used in places where people do not (and should not) enjoy a reasonable expectation of privacy (e.g. airports, border crossings). Second, the cameras equipped with FRT must be marked to assure the public that they are not being surveilled in places that they should have a reasonable expectation of privacy. Third, FRTs should be restricted to finding serious criminals (e.g. terrorists). Fourth, the state should not use third-party companies that violate the first three conditions during the creation or use of its service. And fifth, third-party companies should not be able to access or read the sensitive data collected by the state. With these conditions satisfied, given the effectiveness of FRT, the state can harness FRT's power to counter-terrorism.

The third part, *Technologies that Extend the Reach and Power of the State: Surveillance* then moves to the development and use of surveillance technologies, and how governments seek to justify wider surveillance programs by reference to counter-terrorism efforts. In this part, the authors look at surveillance technologies to show how these technologies when used as part of wider CT programs can make the state much more powerful. John Hardy looks at the general ethical issues that arise when the state engages in surveillance that is persistent, involves pattern-of-life analysis, and activity-based surveillance. Michael Clarke then explores the way that China has used surveillance technologies as part of a "preventative" counter-terrorism campaign in the Xingjian region of China.

John Hardy's chapter "The Rise of the Modern Intelligence State" argues that the rise of the formal surveillance state in the early twenty-first century was precipitated by political impetus to empower security and intelligence organisations to perform a broad range of counter-terrorism functions. Ethical debates about the implications of the security intelligence reach of modern states have focused on balancing individual rights, liberties, and privacy against the security of the state. Meanwhile, the surveillance state has rapidly evolved into an intelligence state, capable not only of pervasive data collection, but also of analytical modelling which expands existing boundaries of surveillance. Existing concerns about the ethical collection and use of surveillance data are compounded by three emergent capabilities of the modern intelligence state: persistent data surveillance, pattern-of-life analysis, and activity-based intelligence. The ethical implications of counter-terrorism intelligence extend beyond the collection and use of data to the application of predictive modelling to dehumanised patterns of behaviour. The chapter shows that this process has the potential to redefine the boundaries of the person, particularly by blurring the distinction between thoughts and actions which threaten the state.

Moving to a particular instance of the surveillance state, Michael Clarke explores the ways that the Chinese government has actively integrated "preventative" counter-terrorism policies that uses new surveillance technologies, particular discourses of the "global war on terrorism" with the ideology of the Chinese Communist Party (CCP) in order to negate the very possibility of "terrorism". The chapter argues that

the contemporary situation in the Xinjiang Uyghur Autonomous Region (XUAR) represents not only the mass repression of an ethnic and religious minority by an authoritarian regime but also an example of the dystopian potentialities of ostensibly “neutral” technologies.

Contrasting surveillance technologies, the fourth part, *The Ethical of Technologies that Limit State Power: Encryption Technologies*, details and expands the ways that encryption technologies can be used to limit state power. In part as a counter-weight to surveillance technologies, encryption technologies offer ways for people to avoid certain state surveillance. Seumas Miller and Terry Bossomaier present a discussion of how encryption technologies work and what the ethical implications of such technologies are. Kevin Macnish then presents an ethical case in favour of encryption.

Starting with the recognition that encryption is obviously a good thing since it protects privacy, but potentially problematic as it might unreasonably impede legitimate counter-terrorism operations, Seumas Miller and Terry Bossomaier explore the technology of encryption technologies. The chapter begins with a general overview of core ethical values relating to encryption and information technologies; privacy, confidentiality, autonomy, and secrecy. It then goes on to show how encryption technologies function. This then allows the final argument of the chapter, a discussion of the privacy rights and security needs in relation to encryption in the overall context of the counter-terrorism policies of liberal democratic states.

Kevin Macnish’s contribution looks at end-to-end encryption, a relatively common technology, that has become even more widespread on mobile phones operating over the Internet. This has provided tools for terrorists to plan activities that lead directly to the deaths of innocent civilians. At the same time, it has also been used by dissidents challenging totalitarian regimes and holding liberal democracies to account. The chapter argues that while terrorist use of such encryption may render that encryption unjustifiable within a liberal democracy, within an international context the protection that it provides to those seeking to establish law-abiding democracies is too great to be ignored.

The fifth part, *Responding to Terrorism in Cyberspace: Extremism Online*, closes the collection out by looking at how the online environment has changed terrorism and what can be done in the name of counter-terrorism. Alastair Reed and Adam Henschke start this part by looking at the ethical issues around who gets to decide to remove terrorists and other political extremists from online environments. Kosta Lucas and Daniel Baldino complete the collection with an examination of the ways that online manifestos can be treated.

A fundamental challenge to modern liberal democracies is how they balance the capacity for free public communication with the need to curtail terrorist use of social media, in a context where this social media dominates people’s lives, public discourse, and even modern politics. Alastair Reed and Adam Henschke ask who should regulate extremist content online. Rather than questions of how this should be done, or what material is relevant, this chapter asks questions of *who* gets to make these decisions and *why*? This chapter suggests that part of the problem with answering “who should regulate extremist content online?” is that there are different aspects to how that content is being regulated. By reflecting on what sorts of institutions and services

are being provided, we can suggest a more nuanced and collaborative approach to the regulation of online content.

Finally, Kosta Lucas and Daniel Baldino take a particular element of online political extremism, to explore identity construction and the usefulness of analysing terrorist manifestos through a narrative framework, with a view to demonstrating that manifestos can be understood as a script to a violent performance (the terrorist act) in the theatre of terrorism (the digital world). The chapter unpacks the dynamic of identity fusion and a specific online terrorist manifesto that coupled with an activist extremist agenda while seeking, in part, to exploit the media in a national security context. The way that this online material is treated has further ethical importance. Media coverage of mass shooters rewards them by making them famous and delivers a clear incentive for future offenders to attack. Instead, the authors argue that if the media modifies how they cover mass shooters, such anticipated changes might be able to deny offenders the personal attention they seek in their quest for significance and help to deter some future perpetrators from normalising violent behaviour.

As with all such projects, there are no simple answers. Moreover, the contributors bring a range of different tools and approaches to these issues, and there is no common consensus on how technologies ought to be used or controlled in the fight against terrorism. This is in part a fact of debates about counter-terrorism, and about technologies, and in part a deliberate feature of the book. These areas are broad, deep, and navigating them is a complex and challenging enterprise. However, there are common threads through the debates—not only must we grapple with terrorism as it evolves, we must also recognise and wrestle with the roles that technologies are playing in the fight against violent extremism. The challenges are considerable, but together we will forge a path to push back the frontiers of counter-terrorism.

## References

1. Gunasingham A (2019) Sri Lanka attacks: an analysis of the aftermath. *Count Terror Trend Analys* 11(6): 8–13
2. Laxman S, Kessler B (2019) Sri Lanka bombings were retaliation for Christchurch shooting, defense minister says NBC News, April 23, 2019. <https://www.nbcnews.com/news/world/sri-lanka-bombing-was-retaliation-christchurch-shooting-defense-minister-says-n997391>. Accessed 27 Jul 2021
3. Ingber S (2019) Global effort begins to stop social media from spreading terrorism. NPR, April 29, 2019. <https://www.npr.org/2019/04/24/716712161/global-effort-begins-to-stop-social-media-from-spreading-terrorism>. Accessed 27 Jul 2021

# Contents

<b>Technology as Terrorism: Police Control Technologies and Drone Warfare</b> .....	1
Jessica Wolfendale	
<b>On the Moral Significance of Narrative, Imagery, and Social Signalling in Counterterrorism Targeted Killing Operations</b> .....	23
Michael Robillard	
<b>Sunlight Glinting on Clouds: Deception and Autonomous Weapons Systems</b> .....	35
Amanda Sharkey and Noel Sharkey	
<b>Weapons of Mass Destruction—Conceptual and Ethical Issues with Regard to terrorism</b> .....	49
Jonas Feltes	
<b>Terrorism and the Internet of Things: Cyber-Terrorism as an Emergent Threat</b> .....	71
Adam Henschke	
<b>Facial Recognition for Counter-Terrorism: Neither a Ban Nor a Free-for-All</b> .....	89
Scott Robbins	
<b>The Rise of the Modern Intelligence State</b> .....	105
John Hardy	
<b>“No Cracks, no Blind Spots, no Gaps”: Technologically-Enabled “Preventative” Counterterrorism and Mass Repression in Xinjiang, China</b> .....	121
Michael Clarke	
<b>Privacy, Encryption and Counter-Terrorism</b> .....	139
Seumas Miller and Terry Bossomaier	

**An End to Encryption? Surveillance and Proportionality  
in the Crypto-Wars** ..... 155  
Kevin Macnish

**Who Should Regulate Extremist Content Online?** ..... 175  
Alastair Reed and Adam Henschke

**White Knights, Black Armour, Digital Worlds: Exploring  
the Efficacy of Analysing Online Manifestos of Terrorist Actors  
in the Counter Terrorism Landscape** ..... 199  
Kosta Lucas and Daniel Baldino

# Technology as Terrorism: Police Control Technologies and Drone Warfare



Jessica Wolfendale

**Abstract** Debates about terrorism and technology often focus on the potential uses of technology by non-state terrorist actors and by states as forms of counterterrorism. Yet, little has been written about how technology shapes how we think about terrorism. In this Chapter I argue that technology, and the language we use to talk about technology, constrains and shapes our moral understanding of the nature, scope, and impact of terrorism, particularly in relation to state terrorism. After exploring the ways in which technology shapes moral thinking, I use two case studies to demonstrate how technology simultaneously hides and enables terrorist forms of state violence: police control technologies and Unmanned Aerial Vehicles (UAVs), or drones. In both these cases, I argue that features of these technologies, combined with a narrative of precision and efficiency, masks the terrorist nature of the violence that these practices inflict and reinforces the moral exclusion of those against whom these technologies are deployed. In conclusion, I propose that identifying acts of terrorism requires a focus on the impact of technologies of violence (whether they are “high tech” or not) on those most affected, regardless of whether users of these technologies conceive of their actions as terrorist.

The topic of this volume is terrorism and technology. Typically, discussions about the relationship between terrorism and technology focus on how new technologies, such as drones [21, 51], artificial intelligence [54], social media [18], and surveillance technologies could be used either as a means of *fighting* terrorism or as a *method* of terrorism [16].

---

This paper benefited greatly from comments from the editors of this volume, Adam Henschke and Scott Robbins. I would also like to thank Risa Brooks, Nicholas Evans, Theresa Tobin, Anthony Peressini, and the faculty and graduate students at the Marquette University Philosophy Department’s Weekly Seminar, for their helpful feedback and suggestions.

---

J. Wolfendale (✉)  
Marquette University, Milwaukee, USA  
e-mail: [Jessica.wolfendale@marquette.edu](mailto:Jessica.wolfendale@marquette.edu)



Few authors, however, recognise how technology shapes and reflects the moral framework through which we think about terrorism, terrorists, and the victims of terrorism—particularly in relation to state terrorism. Instead, the standard view is that “what is good or bad about [technology] is not technologies themselves but the ends to which they are put” ([31], 72). In this chapter I argue that technologies of violence are not simply neutral objects that may be used for good or bad purposes. Instead, the design of these technologies, the contexts in which they are deployed, and the narratives surrounding their use reflect and reinforce biases and frame and limit moral decision-making regarding when and against whom technologies are used. Thus, these technologies profoundly impact our moral understanding of the nature and justification of different forms of violence. Section 1 outlines how both the concept of technology and technological artefacts themselves create and embody normative associations and values that shape the moral landscape of their use. In Sects. 2 and 3, I apply David Rodin’s moral definition of terrorism to the case studies of police control technologies and drone warfare. I argue that police control technologies, including riot control technologies, stun guns, and tasers, function as a *terrorist display* that reflects and reinforces the long-standing and deeply entrenched association of criminality with blackness and thus play a crucial “signifying role” in delineating who may be harmed, who is a threat, and who is to be protected. In Sect. 3, I argue that the US drone program is also a form of terrorism. However, the nature of drone technology, and the accompanying narrative that frames drones as weapons of precision and discrimination, masks the terrorist impact of drone warfare on those subjected to it and contributes to the illusion that drone warfare is objective, precise, unbiased, and even inherently moral.

In both cases, I show how the narrative of technologies of violence as neutral tools masks the terrorist nature of certain kinds of state violence and obscures the power dynamics inherent in that narrative. As will become clear, the view that these technologies are morally neutral or even benign reflects the privileged stance of users of these technologies. From the perspective of those who are subjected to these technologies, they are far from morally neutral. Thus, as I argue in the conclusion, identifying acts as terrorist requires focusing on the impact of those acts (whether they are “high-tech” or not) on those most affected, regardless of whether those involved in producing these effects conceive of their actions as terrorist. Scholars writing on terrorism and technology must acknowledge that the development and use of technologies of violence encodes and reinforces normative judgments about terrorism, the moral status of victims of terrorism, and moral responsibility for terrorism.

## 1 The Concept of Technology

We could define “technology” simply as any human made artefact, including everything from basic tools, “specific devices and inventions,” to “complex sociotechnological systems” ([39], 547). But if that is all we mean by “technology,” there is no reason to think that the relationship between technology and terrorism poses any

unique ethical questions: *of course* terrorists use technology (guns, planes, mobile phones, bombs, and so forth) to achieve their goals, to varying degrees of success, and *of course* technology can be employed to fight terrorism. But this way of thinking about the relationship between technology and terrorism ignores the fact that the term “technology” involves a range of concepts and associations that are not always made explicit, but that shape our moral thinking in important ways.

### 1.1 *Technology and Moral Mediation*

It is a mistake to see technologies as inert objects with which we interact with the world. Instead, as Peter-Paul Verbeek argues, technologies “give shape to what we do and how we experience the world. And in doing so they actively contribute to the ways we live our lives” ([56], 1). Technologies “mediate moral decisions and help to attribute responsibilities and instil norms” ([56], 2).

This process occurs along several dimensions. Firstly, from when it first gained widespread usage in the late nineteenth century, the concept of technology was associated with the idea of moral and social progress ([30], 969). This is particularly true in relation to technologies of state violence. To illustrate, in the US, each time a new technology of execution (electric chair, gas chamber, lethal injection) was introduced, it was heralded as offering not only a more *efficient* means of killing, but a more *humane* means of killing, thereby conflating technological capacity with moral values. For example, one newspaper described the electric chair as providing a death that was “less painful and *more dignified*” ([26], 4, emphasis added). Another claimed that “science has devised a much more effective and *decent* way of putting to death” ([26], 12, emphasis added). Similar statements were made about the gas chamber and lethal injection. Yet, in each case the supposed humanity of the new technology was undermined by the botched executions and visible suffering that occurred almost as soon as the technology was put into use, leading to a further (futile) search for a technological solution to the problem of capital punishment ([26], 22)—a search that obscures the irresolvable moral tension in the very concept of a humane execution. As we shall see, a similar moral tension, and the use of a narrative that conflates efficiency with moral progress, also underlies the search for technological solutions to police brutality, and in the development and use of drones.

The association between technological development and moral and social progress also plays out in the distinction between “high-tech” and “low-tech.” “High-tech” is associated with civilization and progress, whereas “low-tech” suggests primitive societies and backward moral thinking. As Phillip McReynolds argues in his discussion of the discrepancy between Al Qaeda’s low-tech terrorism and the high-tech counterterrorism response of the United States,

the low technology of terrorism [suicide bombs, box cutters, and so forth] bears the marks of a lack of respect for human life in general, for individualism, and for freedom whereas high technology as located within an ideology of progress is understood of leading directly to a greater respect for human life, individuality, and freedom ... the notion of high-tech violence

as opposed to the more direct, low-tech variety carries as sense of moral superiority. ([31], 82–83)<sup>1</sup>

Secondly, technology organises “situations of choice and suggest[s] the choice that should be made” ([56], 5). As Bruno Latour explains, technology can “authorise, make possible, encourage, make available, allow, suggest, influence, hinder, prohibit, and so on.” ([1], 104). Different technologies amplify some aspects of the world and reduce the prominence of others, and thereby “direct” or “organise” our perceptions in particular ways ([56], 11). This has significant, but often underappreciated, moral implications. For example, the mere *availability* of a technology may be viewed as a *moral* reason for selecting it, as occurred when the Dallas Police Department used a bomb-disposal robot carrying C-4 explosives to kill a man who had shot five officers. In defending this action, Police Chief David Brown stated that “*We had no choice*, in my mind, but to use all tools necessary” ([42], 281, emphasis added). The availability of the robot thereby played a role in “directing ... moral deliberations” ([42], 281) and was “influential in justifying such extreme means” ([42], 285). Once a technology is utilised in this way, further use of the technology rapidly becomes normalised and justified and diverts attention away from other possible courses of action: “legitimizing the use of a technology is linked to its naturalization” ([36], 65). Lorna Rhodes makes this point in her discussion of the technology of solitary confinement: “once the option of isolation exists, it tends to be normalized as a ‘common sense’ fix for inadequate mental health care, overcrowding, and failure to adequately protect prisoners in the general population.” ([39], 551).

Thus, the choice of technology shapes moral decision-making in ways that can lead to a conflation between moral concepts such as justification and non-moral concepts such as efficiency. As Elke Schwarz explains, the “moral significance of choosing technological means might make some means that are not necessarily justified *seem* justified; it might make means that are not absolutely necessary *seem* necessary, and it might make technological tools that for whatever reason appear to be the most attractive option in a collection of available options seem like the *only* option” ([42], 284–85).<sup>2</sup>

---

<sup>1</sup> McReynolds attributes this to the ways in which high-tech weapons, such as drones and long-range missiles, make killing seem “less violent ... the more direct connection to it [violence] that accompanies low-tech violence tends to reflect poorly on the human and moral status of the person who carries it out.” ([31], 83). This distinction is also likely part of the reason why “high-tech” violence, such as that inflicted by drone warfare (discussed in Sect. 3), is less likely to be described as terrorism. I thank Risa Brooks for suggesting this point.

<sup>2</sup> Schwarz makes this point in relation to the choices of technology in thought experiments to do with justified killing and liability to harm. For example, in her discussion of Gerhard Øverland’s thought experiment involving the use of a flamethrower in self-defense that threatens the lives of others nearby, she writes: “Øverland posits that the permissibility of Mary using her flamethrower and killing the occupants of the house depends on how many people would die and how many would be saved. In this case, the cost of the destructive range of the technology at hand is assigned to the people in the house, who become ‘moral obstacles’, despite the fact that the availability of the flamethrower as a specific means of action is entirely invented by the thought experiment” ([42], 284).

## 1.2 *Technology and Bias*

Technologies often embody and reinforce the moral, social, and political norms and biases of those who create and use them. One obvious way this occurs is when an otherwise “neutral” technology is deployed in ways that disproportionately harm members of a certain group as, for example, when police control technologies such as tasers and stun guns are used disproportionately against persons of colour. But biases and norms can also be literally “built in” to technological systems in ways that can cause disproportionate harm to members of minorities and other stigmatized groups.

Algorithms offer one example of bias in the design and use of technology. As Schwarz explains, “how an algorithm functions and how it is trained reflects the values and principles of its intended uses and its designers ... They regularly reflect the aims and intentions of their makers and normalize their positions and priorities (values)” ([42], 292). For example, studies on facial recognition technologies in the context of law enforcement have found that these technologies reflect and reinforce racial bias. Ruha Benjamin describes the scale of this “default discrimination”: “At every stage of the process—from policing, sentencing, and imprisonment to parole—automated risk assessments are employed to determine people’s likelihood of committing a crime.” Yet, multiple studies have found that these automated processes are “remarkably unreliable in forecasting violent crime” ([5], 81). The impact of this encoded bias can be devastating: “Black people are overrepresented in many of the databases faces are routinely searched against” which means that “Black people are more often stopped, investigated, arrested, incarcerated and sentenced as a consequence of facial recognition technology ... Black people are more likely to be enrolled in face recognition systems, be subject to their processing and misidentified by them” ([4], 326).

The problem of biased algorithms in facial recognition systems is exacerbated by the phenomenon of automation bias [12]. Research demonstrates that humans have an unwarranted belief in the neutrality and accuracy of technological systems: “humans have a tendency to disregard or not search for contradictory information in light of a computer-generated solution that is accepted as correct” ([42], 290). This means that the “results” of facial recognition algorithms (and other biased algorithms) are likely to be assumed to be objectively correct, leading to a vicious cycle that reinforces embedded biases and lends them an unwarranted patina of legitimacy ([12], 2–3).

Kodak’s Shirley card is an example of bias that is literally “built in” to a technological system. The Shirley card was used as a comparison image to ensure that the colours in a printing look “right”. In its original form, the Shirley card featured a white woman with “ivory skin, brown hair, and red lipstick” ([25], 3). But, “[s]ince the model’s white skin was set as the norm, darker skinned people in photographs would be routinely underexposed” ([5], 104). The Shirley card thus both reflected its creators’ racial biases and then continued use of the Shirley card reinforced this bias, calcifying the view that white skin was the ideal aesthetic standard and the standard of “normal” skin tone (see [5], 103–109).

In sum, technologies “mediate moral decisions” ([56], 2), and so shape our moral understanding of our actions by offering (and restricting) choices, reflecting and reinforcing pre-existing biases, and through the development of accompanying narratives that frame new technologies in terms of moral values such as dignity and humanness. As is clear from the example of capital punishment discussed earlier, the narratives that accompany the development and use of new technologies frequently privilege the perspective of users and developers rather than that of those subjected to these technologies. In what follows, I show how this complex dynamic between technology and moral evaluation and decision-making plays out in the context of drone warfare in ways that obscure the impact of drone warfare on those subjected to it—an impact that is, I argue, sufficiently severe to constitute terrorism.

### 1.3 What Is Terrorism?

What do I mean by terrorism? In this chapter, I adopt elements of David Rodin’s *moral definition* of terrorism. A moral definition is “an analysis of the features of acknowledged core instances of terrorism [such as the 9/11 attacks] which merit and explain the moral reaction which most of us have toward them” ([40], 753). Rodin locates the moral opprobrium many of us feel toward terrorism in the fact that core instances of terrorism are characterised by “the use of force against those who should not have forced used against them” ([40], 755). He then defines terrorism as “the deliberate, negligent, or reckless use of force against noncombatants, by state or nonstate actors for ideological ends and in the absence of a substantively just legal process” ([40], 755).<sup>3</sup> The reference to force against noncombatants for ideological ends is consistent with many other definitions of terrorism. Rodin’s inclusion of reckless and negligent acts in his definition is controversial but given that the case studies I discuss involve intentional actions, I will not weigh in on this controversy here.<sup>4</sup> Given this definition, we can now turn to the case of police control technologies.

## 2 Police Control Technologies as Terrorist Display

Police control technologies include devices such as tasers and stun guns, as well as riot control technologies such as tear gas, rubber bullets, and the use of militarised

---

<sup>3</sup> Rodin defines “ideological ends” to “signify a commitment to some systematic and socially directed end beyond the motives of fear, anger, lust and personal enrichment, which are the typical motives of common violent crimes” ([40], 756). The term “noncombatants” is intended to capture the fact that the victims of terrorism are not engaged in activities that would render them liable to the use of force, such as combat. Thus, attacks against military targets can count as terrorism ([40], 757). Reference to the absence of a “substantively just legal process” is intended to distinguish terrorist violence from the use of force accompanying just legal processes ([40], 759–60).

<sup>4</sup> See [60] for a critique of Rodin’s claim that reckless and negligent acts can count as terrorism.

weapons, tactics, and uniforms “that were once the preserve of military units in war zones” ([14], 110). The contexts in which these technologies are used, the class of people against whom they are deployed, and the justifications offered for their use, reveal much about who is perceived as a threat, who is judged liable to be killed and wounded, and who is judged worthy of protection.

## 2.1 Riot Control Technologies

### 2.1.1 The Narrative of Threat

A justificatory narrative of threat and protection is particularly apparent in the use of riot control technologies. This means that the contexts in which riot technologies are *not* used are just as revealing as the contexts in which they are used. For example, in the wake of the killing of George Floyd, Black Lives Matter (BLM) protesters were subjected to tear gas and other “non-lethal weapons” such as rubber bullets and stun grenades, wielded by police and federal forces clad in militarised riot gear, including face shields, external bullet-proof vests, and knee-high boots. In comparison, the armed white protestors who raided the US Capitol building on January 6, 2021, faced police who were not clad in riot technology and who did not engage in substantial force against them [58]. This stark and visible disparity in the use of violent control technologies serves a powerful signifying function: BLM protestors are dangerous but white protestors are not, even when engaged in a violent armed insurrection, the technologies of violence and suppression are *necessary* (and therefore justified) when interacting with BLM protestors, but not when interacting with majority white protestors [37]. Images of the police response to these different groups, replicated in media coverage of the protests, communicates and reinforces, even more effectively than words or political speeches, the criminalisation of blackness<sup>5</sup> and the belief that people of colour (and those who support them) pose such a threat that they may justifiably be harmed or killed. The visual narrative that accompanies the use of these technologies thereby “symbolically excludes the citizens from the state” ([14], 114) and reflects a resurgence of the “escalated force” policy of “a dominant show of force” that governed police responses to anti-war and civil rights protestors in the 1960s (groups also characterised as threats to the state) ([29] 75).

### 2.1.2 Techno-Subjectivity and Moral Mediation

The “techno-subjectivity” ([42], 288) of these technologies (how it feels to deploy and wear them) feeds this narrative of threat and mediates the moral decision-making of

---

<sup>5</sup> As noted in Sect. 1, this narrative is also embedded and reinforced through the design and use of facial recognition technologies.

those who wield them. There is substantial evidence that when police adopt military-style tactics and “start using weapons and equipment that were designed for soldiers in combat” ([14], 109), their perception of their role and their relationship with the community is altered, particularly in relation to communities of colour: “pacifying and defeating the enemy becomes more important than protecting and serving the public” ([14], 110. See also [37]). In the United States, the adoption of military technology also has a measurable impact on incidents of police killings. One study found that “more than twice as many civilians are likely to be killed by police in a county after its material militarization than before” ([14], 111). This risk is not distributed evenly among the community, however: “Risk is highest for black men, who (at current levels of risk) face about a 1 in 1,000 chance of being killed by police over the life course. The average lifetime odds of being killed by police are about 1 in 2,000 for men and about 1 in 33,000 for women ... For young men of color, police use of force is among the leading causes of death” [15].<sup>6</sup> Thus, the deployment of riot control and other militarised technologies reinforces the association of blackness with criminality and directly contributes to the ongoing and pervasive vulnerability of people of colour to violent interactions with criminal justice system. The ready availability of these technologies combined with the contexts in which they are (and are not) deployed thereby creates an ongoing and embedded “feedback loop” that reinforces the belief that people of colour and their supporters represent a dangerous threat. This feedback loop is sustained through at least three mechanisms: the narrative of threat described above, the accompanying media circulation of visual images of riot technologies deployed against people of colour, and the phenomenological impact on police of wielding these technologies.

### 2.1.3 The Terrorist Impact of Riot Technologies

Riot control technologies not only communicate and reinforce the criminalisation of blackness and the moral exclusion of people of colour from the moral and political community; they have concrete traumatic effects that justify the claim that the deployment of these technologies is a form of terrorism. Firstly, the use of these technologies against peaceful protestors communicates a very real threat of physical violence that signifies to those subjected to them that they may be killed or harmed with impunity. Secondly, these technologies cause severe and lasting physical injuries, fear, and ongoing trauma [43]. The fact that these technologies are used disproportionately against people of colour and other groups deemed to be outside the moral and political community (such as anti-war protestors in the 1960s and 1970s) indicates that their use is ideologically driven. The ideological nature of these technologies is further evidenced by the origins of their use: “the so-called non-lethal crowd control weapons that are used to disperse protests today have their origins in colonial policing” [43], where there were used to violently reinforce white

---

<sup>6</sup> There are similarly disproportionate rates of police violence against indigenous Australians compared to non-indigenous Australians [13].

supremacist colonial regimes against resistance. As a scholar of the history of tear gas argues, these technologies (then and now) were “deployed to both physically and psychologically destroy people engaging in resistance” (quoted in [43]). The impact of these technologies and the way these technologies are deployed, therefore, clearly meets Rodin’s definition of terrorism as “the use of force against those who should not have force used against them” that serves an “ideological end” ([40], 753).<sup>7</sup> Give the role of these technologies in creating and sustaining the long-standing and deeply entrenched criminalisation of blackness and the vulnerability of people of colour to police violence, it is not a stretch to say that these technologies are part of a broader system of terrorist control of people of colour. This is also demonstrated by the use and development of tasers and stun guns.

## 2.2 *Tasers and Stun Guns*

### 2.2.1 **The Narrative of Effectiveness and Humaneness**

While the use of riot technologies is accompanied by (and reinforces) a narrative that focuses on threat, the narrative accompanying the development and use of stun guns and tasers by police appeals to the values of humanness and effectiveness, similar to the narrative that accompanied the development of new execution technologies. When tasers were first introduced as police control technologies, for example, they were touted as being “safe, effective alternatives to ... lethal force” ([45], 421) that would solve the ongoing problem of the disproportionate use of excessive (sometimes lethal) force by police against people of colour. (Similar claims have been made about body cameras.) Yet, the problem of excessive force has not in fact diminished [24]. Instead, the availability of tasers (and stun guns) gave police officers an option they did not previously have, and one that was framed in morally positive terms as non-excessive and humane. But, just as describing new execution technologies as humane did not in fact make executions more humane, the framing of tasers as non-excessive did not in fact mitigate police of force.<sup>8</sup>

This illustrates how describing tasers as a technological solution to the problem of excessive police violence implies that the problem of excessive force is a *technological* problem that requires a *technological* solution, and not a problem arising from the longstanding and well documented framework of racism that underpins and structures policing interactions with (and attitudes toward) people of colour in the US [53].

---

<sup>7</sup> Someone might object that violent protesters count as combatants and so these technologies do not target “those who should have force used against them.” However, riot technologies are often used against peaceful protestors and there is little attempt to restrict the use of force to those who act violently. Additionally, the visual communication of the threat of violence is indiscriminate in its impact.

<sup>8</sup> I thank Scott Robbins for suggesting this point.



## 2.2.2 The Terrorist Impact of Tasers and Stun Guns

Those who defend the use of tasers and stun guns may frame them as technologies of non-lethal restraint and control that can (if properly used) “not appear cruel or beneath human dignity” ([38], 157). But the widespread acceptance and normalisation of the use of stun guns and tasers masks the history of these devices in the contexts of torture and animal control, a connection that is apparent to those who are subjected to these devices. From the victims’ perspective, the use of electric control technologies does not signify respect for their dignity, a reduction in force, or a humane method of control. As Lorna Rhodes relates, prisoners in Supermax prisons (where stun guns are used as control mechanisms), “speak of these technologies as particularly degrading both for their extreme intrusion into the body (they cause muscle weakness as well as pain) and for their association with the control of animals” ([39], 556). But, the victims’ experiences of these technologies as degrading, dehumanizing, and torturous is masked by the dominant narrative of efficiency and humaneness that frames their use. Thus, this narrative both reinforces and hides the true function of these technologies and privileges the perspectives of users above that of those who are subjected to them.

The association of tasers and stuns guns with torture (a long-standing method of state terrorism) is also clear from the history of these devices in the context of state torture. As Darius Rejali explains, stun guns and other electric devices are popular in states that use torture because, like other “modern” torture techniques (such as sensory deprivation), they “cause suffering and intimidation without leaving much in the way of embarrassing long-term visible evidence of brutality” ([38], 153). In the context of torture, the use of these technologies is not driven by a concern for human dignity, but by a desire to avoid charges of human rights violations. Given this history, the widespread acceptance and availability of electric control technologies in the context of law enforcement is astonishing. It represents “an incredible sociotechnical achievement, the work of corporations, politicians, and engineers who have woven this technology into the fabric of everyday life, creating instruments, markets, citizens, and consumers” ([38], 154–55). As with riot control technologies, those against whom this technology is wielded (who are disproportionately prisoners and people of colour, and those who threaten the state in other ways) are thus “marked out” as deserving or requiring such violent treatment. The use of these technologies (as with the deployment of riot control technologies) thereby operates as what Rejali calls “a civic marker” ([38], 154) delineating the moral boundaries of civic membership and moral concern through the infliction of instruments associated with terror and torture.

### 2.3 Implications

The above discussion has several implications for understanding the relationship between police control technologies and police use of force. Firstly, any ethical analysis of policing technologies must address how some technologies directly “encode” racial bias (as with facial recognition algorithms). Secondly, such an analysis must also recognise how the contexts in which these technologies are used, and the narratives accompanying their use, shape and constrain the moral decision-making of police officers (and policy makers) in ways that reflect and reinforce an underlying framework of racism. This means that the problem with riot control technologies, tasers, and stun guns is not a problem that can be solved by better training or new policies about the contexts of their application. As we have seen with the failure of body cameras and implicit bias training to reduce rates of police violence against people of colour [24], unless the deeply embedded racist structure of policing in America is confronted and addressed, police technologies will continue to be utilised in ways that reinforce that racist structure and terrorise and threaten the lives and welfare of people of colour. It is for this reason that the “defund the police” movement has gained traction over the last year—a movement that calls for moving state and federal funding and resources from the police and criminal justice system to (for example) social services, public education, mental health services, and affordable housing. This would, it is argued, not only reduce crime rates but increase the safety and wellbeing of all citizens, and particularly people of colour. Such a move is arguably justified not only economically [33] but also because it would also go some way to addressing the underlying issue (one I cannot address in detail here) that terrorist policing practices against people of colour undermine the very basis of the state’s authority to use force against its own citizens in a criminal justice context.<sup>9</sup>

## 3 Drone Warfare

As with the case of police control technologies, the terrorist nature of drone warfare results from the combination of features of drone technology (the capacity for long-term surveillance and the use of algorithmic targeting decisions), the contexts in which drones are deployed, and the impact on those who are subjected to drone surveillance and targeting. This terrorist impact is masked by a narrative that frames the use of drones as morally neutral, even morally good. But whereas the narrative associated with police control technologies emphasised threat protection, control,

---

<sup>9</sup> In many philosophical accounts, the basis for the state’s authority to use force against its own citizens to prevent and punish crime is a social contract model (e.g., see [6]). Thus, if police actions and the criminal justice system threaten rather than protect citizens, this undermines the fundamental basis for the legitimacy of such systems. Just as Adam Henschke and Tim Legrand have argued in relation to counter-terrorism policies, we need to ensure that the technologies being used by police do not in fact run counter to the values that underpin and justify the monopoly of power granted to the state [17]. I thank the editors of this volume for raising this concern.

and humaneness, the narrative that dominates military and political discourse about drones emphasises precision and discrimination.<sup>10</sup> As the Center for Civilians in Conflict reports, “as covert drone strikes by the United States become increasingly frequent and widespread, reliance on the precision capabilities and touted effectiveness of drone technology threatens to obscure the impact on civilians” ([19], 7). This narrative, and the features and context of drone use, thereby serve to “morally mediate” ([56], 2) the use of drones by constraining moral choices around drone use, shaping the moral perception of users, policy makers, and the public about the nature and justification of drone use, and “marking out” the targets of drone attacks as warranting the use of force against them.

This means that the terrorist nature of drone warfare only becomes evident when we shift our focus from the narrative and associated moral framework that dominates discussion of drones to the impact of the drone program on those who are subjected to it. First, however, we need to clarify the current scope of the US drone program.

### 3.1 *The US Drone Program*

The use of drones as a means of killing suspected and known members of Al Qaeda and other terrorist and militant organisations began under the Bush administration, expanded under the Obama administration ([21], 3–4), and expanded further under the Trump administration. According to one report, “As of May 18, 2020, the Trump administration had launched 40 airstrikes in Somalia in 2020 alone.” In contrast, “from 2007 through 2016, the administrations of George W. Bush and Barack Obama conducted 41 airstrikes in Somalia total.” [3]. Additionally, the Trump administration broadened the designation of “battlefields” to include areas of Yemen and Somalia, thereby loosening the restrictions on drone targeting in those areas [3]<sup>11</sup> and simultaneously “removing the reporting requirement for casualties outside of designated battlefields” [3]. This led to a dramatic increase in the numbers of civilian casualties of drone strikes: “In 2019, more Afghan civilians were killed in airstrikes than

---

<sup>10</sup> There is a substantial philosophical literature on the ethics of drones (see, for example, [21, 51]), which I do not have space to discuss here. Ethical issues raised by authors include concerns about the asymmetry of drone warfare [23, 50], the impact of drone warfare on the moral equality of combatants [46], the moral disengagement of drone operators ([44], 371–72), drone operators’ moral responsibility ([48], van der Linden 344), and the effect of drone warfare on conceptions of traditional military virtues [49]. Several authors regard the ethics of drone use as no different from the ethics of any long-range technology [22, 27]. For example, George Lucas argues that, “[a]s with most exotic new technologies, the novelty [of drones] blinds us to the fact that the moral issues involved are entirely familiar and conventional and not appreciably different from those associated with the development of previous and current weapons technology” ([27], 211).

<sup>11</sup> The Obama administration’s Presidential Policy Guidance (PPG) designated looser targeting restrictions for battlefields and tighter ones for nonbattlefields, to allow drones greater freedom in “providing support fire for soldiers in firefights in places such as Afghanistan, while holding tighter restrictions for targeted killing flights in places where the United States did not actively deploy troops on the ground, such as Yemen or Somalia” [3].

at any time since early 2002” ([11], 2). While the Biden Administration has introduced some restrictions on drone use, including temporarily suspending the use of drones outside war zones [41], it remains unclear what the scope of these changes will be or how, for example, targeting decisions within war zones will be made. This lack of clarity became evident with the release of the Pentagon’s investigation into the August 29, 2021, drone strike that killed 10 civilians (including seven children) in Afghanistan, that found that no laws were broken but that “communication breakdowns” occurred [47]. While much remains unknown about this strike, and the long-term intentions of the Biden administration regarding the use of drones, it seems clear that the drone program will be ongoing and there will continue to be little transparency about the impact of drone warfare on those most affected by it.

## 3.2 *Drone Warfare as Terrorism*

### 3.2.1 The Narrative of Precision and Discrimination

From their introduction drones have been heralded as “precision weapons” that allow war to be conducted in a more humane way:

US intelligence officials tout the drone platform as enabling the most precise and humane targeting program in the history of warfare. President Obama has described drone strikes as “precise, precision strikes against al-Qaeda and their affiliates.” Leon Panetta, Secretary of Defense, has emphasized that drones are “one of the most precise weapons we have in our arsenal,” and counterterrorism adviser John Brennan has referred to the “exceptional proficiency, precision of the capabilities we’ve been able to develop.” ([19], 35)

As a result of this narrative, “public concerns with civilian casualties in targeted killing campaigns—concerns that are generally weak or even nonexistent to begin with—are put to rest” ([55], 335).<sup>12</sup> As we saw with the language that accompanied the development of new execution technologies, this emphasis on precision conflates a *technological* value with a *moral* value (“humaneness” or “dignity”). The view that the technical capacity of drones to distinguish between targets is also a moral capacity is shared by some philosophers. Bradley Strawser, for example, argues that a drone’s capacity to discriminate between targets combined with the fact that drone use reduces the risk to the operator to essentially zero means that “we are morally required to use drones over ... manned aircraft to prevent exposing pilots to unnecessary risk” ([52], 18).

However, conflating drones’ *technical* capacity for precision targeting with the *moral* distinction between combatants and noncombatants not only sustains and

---

<sup>12</sup> It is extremely difficult to know the precise number of civilians who have been killed by drone strikes. This is a result of a combination of factors, including difficult terrain that makes on-the-ground verification impossible, and the ways in which the category of “militant” is sometimes used to describe any “military-aged male” killed in a strike [7]. However, my argument for the terrorist nature of drone warfare does not rest only on the numbers of civilians who are killed.

reinforces an unfounded complacency about the morality of drone strikes but also obscures the reality of who is targeted by drones and for what reasons. As Harry van der Linden notes, “precision in finding and hitting the target does not imply that there is precision in the *selection* of the target” ([55], 336, emphasis in original). John Kaag and Sarah Krepps make the same point: “The distinction between militants and non-combatants ... is a normative one that machines cannot make” ([21], 134). Put simply, we cannot assume that the categories of combatant and noncombatant are either clearly defined or justly applied by drone operators and/or political and military decision-makers in the drone program. In fact, we have good reason to doubt that this is the case. For example, claims by US officials in the Obama administration that drones strikes caused very few civilian casualties ([7], 31) were complicated by the fact that these assertions were based on “a narrowed definition of ‘civilian,’ and the presumption that, unless proven otherwise, individuals killed in strikes are militants” ([7], 32). As I argue below, the assumption that the targets of drone strikes are chosen based on clear and justly applied categories of combatant and noncombatant is extremely problematic.

### 3.2.2 Bias and the Moral Mediation of Drone Technology

In Sect. 1.3, I explained how bias can be “built in” and reinforced by technology in multiple ways, from the design of algorithms and the physical features of technologies, to choices about when and against whom technologies are deployed. These forms of bias can become entrenched because of the normalising and self-justifying effects of repeated use of a technology in a specific context against specific groups of people, combined with the phenomenon of automation bias—the tendency of users and designers of technologies to assume that the “answers” provided by technological systems are both objective and correct [12]. In the cases of drones, bias is evident both in the algorithms that are used to select the targets of drone strikes and in how the class of acceptable targets (who are almost exclusively non-white people) has expanded far beyond any plausible definition of “combatant.” This bias is most apparent in the use of drones for signature strikes.

Unlike targeted strikes, where the identity of the target is confirmed before a strike is permitted, signature strikes may be initiated on the basis of perceived patterns of suspicious behaviour: “Signatures may encompass a wide range of people: men carrying weapons; men in militant compounds; individuals in convoys of vehicles that bear the characteristics of al-Qaeda or Taliban leaders on the run, as well as ‘signatures’ of al-Qaeda activity based on operatives’ vehicles, facilities, communications equipment, and patterns of behavior” ([7], 33). But the value of signature identifications depends on a host of normative and culturally biased assumptions about what counts as “suspicious” behaviour.<sup>13</sup> As Elke Schwarz argues, the use

---

<sup>13</sup> As related in *The Civilian Impact of Drones*, “As one Yemeni official said, ‘Every Yemeni is armed...so how can they differentiate between suspected militants and armed Yemenis?’” ([7], 33). It seems that such bias was present in the events leading up to the August 29, 2021 strike as

of algorithms to determine the targets of signature strikes “summon[s] the perception that patterns of normality (benign) and abnormality (malign) can be clearly identified” ([42], 288).

However, as we saw with the use of facial recognition algorithms in law enforcement, the success of such algorithms in correctly ascertaining and predicting malign intent is highly questionable.<sup>14</sup> Yet, when combined with the phenomenon of automation bias, the “output” of the algorithms used for signature strikes is unlikely to be questioned. This then further reinforces the belief that the mere presence of “suspicious” behaviour (defined based on culturally biased assumptions) provides sufficient evidence of malign intent to justify the use of lethal force. The decision to resort to lethal force is then framed as the “right” or most “logical” response to the perceived threat because “the drone can only execute a limited range of actions vis-à-vis a suspect (survey, pursue or kill). A suspect cannot surrender or persuade the technology of their non-liability to harm” ([42], 288). Thus, the combination of embedded bias in targeting algorithms and the limits of drone technology constrains and shapes the moral choices of users and alters the justificatory framework used to assess the morality of drone warfare. These moral choices and justificatory framework are then normalised via further use of drones combined with the narrative of precision and discrimination discussed above. In particular, this process reinforces and normalises the view that a person may be killed not because they are currently engaged in combat or are known to be part of a militant group, but merely because their behaviour *resembles* that of someone who *might* be a future threat. The technology translates “probable associations between people or objects into actionable security decisions” ([2], 52). This represents an extraordinary broadening of the concept of a combatant that has devastating consequences:

US experiences in Afghanistan illustrate the risks of targeting with limited cultural and contextual awareness. On February 21, 2010, a large group of men set out to travel in convoy. They had various destinations, but as they had to pass through the insurgent stronghold of Uruzgan province, they decided to travel together so that if one vehicle broke down, the others could help. From the surveillance of a Predator, US forces came to believe that the group was Taliban. As described by an Army officer who was involved: “We all had it in our head, ‘Hey, why do you have 20 military age males at 5 a.m. collecting each other?’... There can be only one reason, and that’s because we’ve put [US troops] in the area.” The US forces proceeded to interpret the unfolding events in accordance with their belief that the convoy was full of insurgents. Evidence of the presence of children became evidence of “adolescents,” unconfirmed suspicions of the presence of weapons turned into an assumption of their presence. The US fired on the convoy, killing 23 people. ([7], 47)

---

well. Gen. Sami D. Said, speaking after the Pentagon’s investigation of the case, “blamed a series of assumptions, made over the course of eight hours as U.S. officials tracked a white Toyota Corolla through Kabul, for causing what he called “confirmation bias,” leading to the Aug. 29 strike.”

<sup>14</sup> It is also very difficult to know how the veracity of signature strikes could be ascertained, not only because the targets are not known by name (but are chosen merely based on supposedly suspicious behavior), but also because of the factors that impede identification of drone victims in general, noted in footnote 12.

A similar process of assumptions about “suspicious” behaviour creating and then reinforcing the belief that a strike was necessary and that the targets were terrorists seemed to have also occurred in the August 29, 2021 drone strike. In the wake of the Pentagon’s investigation into the strike, the Air Force’s inspector general, Lt. Gen. Sami D. Said, “blamed a series of assumptions, made over the course of eight hours as U.S. officials tracked a white Toyota Corolla through Kabul, for causing what he called “confirmation bias”” [9]. The relatively high level of media coverage of the August 29, 2021 strike illustrates how little coverage there has been about previous cases of civilian deaths from drones. The killing of people based purely on biased and highly unreliable computer-predicted assumptions about the meaning of their behaviour is taken for granted to such an extent that it is rarely deemed worthy of comment. Indeed, the combination of the narrative of discrimination, drone technology, and the processes of moral mediation discussed above has created a situation where the ongoing killing and maiming of non-white people based on biased assumptions of threat has come to seem both morally acceptable and even necessary.<sup>15</sup> As Elke Schwartz explains, “set against a background where the instrument is characterised as inherently wise, the technology gives an air of dispassionate professionalism and a sense of moral certainty to the messy business of war” ([42], 88). This “moral certainty” is sustained and reinforced by the “high-tech” nature of drone operations and the narrative of precision and efficiency described above and effectively masks the reality of the terrorist impact of drones on the victims.

### 3.2.3 The Terrorist Impact of Drone Warfare

As discussed above, the use of signature strikes significantly increases the risk that noncombatants will be killed and wounded and reinforces the view that merely suspicious behaviour warrants the use of deadly force. But this is only one reason why the current drone program was, and likely remains, terrorist. Even if drone strikes only killed known targets,<sup>16</sup> the impact of living under drone surveillance affects *everyone* in the area under surveillance, whether they are targets or not. Unlike other long-range weapons systems, “only drone killing involves detailed surveillance of the target, including post-strike observation” ([55], 345–46).

The *Civilian Impact of Drones* report produced by the Center for Civilians in Combat and the Columbia Law School Human Rights Clinic outlines the traumatic effects of living under drone surveillance.<sup>17</sup> Firstly, drones engaged in surveillance

---

<sup>15</sup> The killing of non-white *known* targets is also largely unquestioned and normalised, even when the targets are chosen for the purposes of punishment and retaliation (which are not legitimate reasons for killing in just war theory), as was the case with the recent retaliatory drone strike in Syria [10].

<sup>16</sup> I am leaving aside the important question of whether drone strikes against known targets are permissible. My argument is that even if they are, this does not mitigate the terrorist impact of drone warfare.

<sup>17</sup> The report *Living Under Drones*, produced by Stanford University and NYU, also details the psychological trauma caused by living under drones ([7], 55–99).



are constantly visible and audible to all those being surveilled, regardless of whether they are targets or not. As van der Linden describes, “[e]veryone is swept up in the surveillance, and living under drones is living under constant fear since, even as a civilian, one may at given moment be wounded or killed” ([55], 351–52). In an important sense, then, “drones are in their psychological impact indiscriminate weapons” ([55], 351). This psychological impact is extremely traumatic. An interviewer for a UK charity spoke to a Pakistani man who “saw 10 or 15 [drones] every day. And he was saying at night-time, it was making him crazy, because he couldn’t sleep. All he was thinking about at home was whether everyone was okay. I could see it in his face. He looked absolutely terrified” ([7], 24).

Because of the secrecy of the drone program, those living under drone surveillance may have no idea who is being targeted or the basis on which targets are selected. This uncertainty compounds this constant fear that one (and one’s family and loved ones) may be killed or wounded:

With US targeting criteria classified, civilians in Pakistan, Yemen, and Somalia do not know when, where, or against whom a drone will strike. The US policy of ‘signature strikes’ ... substantially compounds the constant fear that a family member will be unexpectedly and suddenly killed. A civilian carrying a gun, which is a cultural norm in parts of Pakistan, does not know if such behavior will get him killed by a drone. ([7], 29)

This perfectly illustrates the “intrusion of fear into everyday life” that Michael Walzer identifies as one of the key moral harms of terrorism [57].<sup>18</sup> The terrorism of drone warfare thus lies not only in the direct physical violence inflicted by drone attacks (which may often kill and maim noncombatants) but also in how drone warfare creates and promulgates a constant, indiscriminate, and terrifying fear of attack.

Compounding the harm of drone warfare is the fact that those who survive a drone attack will often have no way of discovering who attacked them. They are denied access to the norms of accountability: “For victims in particular, there is no one to recognize, apologize for, or explain their sorrow; for communities living under the constant watch of surveillance drones, there is no one to hold accountable for their fear” ([19], 24).

Despite the devastating toll of drone surveillance on those subjected to it, philosophers writing on drones rarely discuss or even mention this aspect of drone warfare.<sup>19</sup>

---

<sup>18</sup> Walzer is not using this term in a discussion of the drone program, however. I do not think he would agree with my characterisation of the drone program as terrorist.

<sup>19</sup> Harry van der Linden is one of the few philosophers who does consider the victims’ perspective. While he does not describe the drone program as terrorist, he argues that the “deadly surveillance” of drone warfare explains why drones may be “inherently immoral” ([55], 345). For van der Linden, drone surveillance is immoral because drone strikes kill people when they are engaged in their ordinary lives—at funerals, while they are under medical care, and in their homes—and this further erodes the distinction between combatant and noncombatant and between battlefield and nonbattlefield. He writes, “operators often become familiar with the target as a person, watch his everyday life, his home, even his family. Thus it seems that a person is killed rather than a combatant or individual engaged in hostile action” ([55], 348). For example, he quotes drone pilot Colonel William Tart saying, “We watch people for months. We see them playing with their dogs or doing their laundry. We know the patterns like we know our neighbors’ patterns. We even go to their funerals” ([55], 350).



For example, Mark Coeckelbergh explores the impact of conducting long-term surveillance on drone pilots' ability to empathise with surveillance subjects [8] but doesn't mention the experience of those living under surveillance. This focus on the experiences of drone operators rather than on the experiences of those who are subjected to the drone program is typical of most philosophical discussions of this topic. It is also characteristic of media depictions of drone warfare. Whereas media depictions of police riot control technologies make visible and reinforce the criminalisation of blackness that underpins the use of those technologies, media depictions of drones almost always show the aircraft themselves, or the cockpits. It is extremely rare that media images show the impact of drone attacks. Thus, viewers are constantly reminded of the technological "marvel" of these weapons and rarely confronted with what these weapons do to the people killed and wounded by them and those who must live under the near-constant threat of attack. This focus on drone pilots and drone technology further prioritises the perspective of users over those of victims of these technologies.<sup>20</sup>

In sum, the US drone program meets Rodin's definition of terrorism because it is an ideologically driven<sup>21</sup> program that inflicts extreme and ongoing psychological and physical trauma on *all* those who are subjected to drone targeting and surveillance, whether they are the intended targets or not.<sup>22</sup> In the absence of clear evidence that the targeting decisions and technological features of the US drone program will substantially change in the foreseeable future, the drone program will likely continue to be a terrorist program under the Biden administration.<sup>23</sup>

---

<sup>20</sup> I thank Desiree Valentine for raising this issue.

<sup>21</sup> It is ideologically driven because it is in service of US foreign policy, which is a "systematic and socially driven end" ([40], 756).

<sup>22</sup> It might be objected that this is true of war in general, given that many of today's wars do not adhere to clear lines between battlefield and nonbattlefield, and between combatant and noncombatant. If that is so, then I would agree that we should consider such wars as inherently terroristic. I thank Scott Robbins for raising this possibility.

<sup>23</sup> Some might argue that, even if the drone war constitutes terrorism, the war may still be justified because of the continuing threat posed by Islamic terrorism. While I do not have space here to address the long-standing debate about whether terrorism can be justified (see [35] for an overview of the debate), this argument fails to justify the drone war because Islamic terrorism does not now (and arguably never did—see [20, 32, 59] pose the kind of existential threat that would be necessary to justify a resort to terrorism, (see, for example, [34])). Indeed, white supremacist terrorism arguably poses a greater threat to the lives of US citizens than Islamic terrorism. For example, the F.B.I. director Christopher Wray described "racially motivated violent extremism" as a "national threat priority" equal to the threat from the Islamic State, and when the New Jersey Office of Homeland Security and Preparedness issued its terrorism threat assessment for 2020, "[t]he threat level from violent, homegrown extremists, and specifically white supremacists, was marked in red as the top category: 'High.' The threat from the Islamic State, Al Qaeda and their ilk was demoted to third, in green: 'Low.'" [28].

## 4 Conclusion: Terrorism from the Victim's Point of View

Terrorism, as characterised by Rodin as the use of force against those who should not have force used against them, is a morally abhorrent practice. The moral abhorrence of terrorism is shared by most writers on terrorism, including myself, and is reflected in common usages of the term. Yet, in this Chapter I have argued that two forms of state violence—police control technologies and drone warfare—are forms of terrorism, despite rarely if ever being described by that word. I have shown that the terrorist nature of these forms of violence is hidden by features of the technologies themselves, the subjectivity of their use, and by the dominant narratives accompanying them. The narratives of efficiency, neutrality, and precision masquerade as moral values and serve to normalise and justify these forms of violence and mark out those subjected to them as deserving of violent treatment. To understand the terrorist nature of these practices, therefore, we must reject the point of view that treats technologies of violence as neutral objects and shift our focus to the experiences of those who are subjected to them. This should always be our starting point when asking whether a practice is a form of terrorism. Such a victim-centred approach to terrorism would destabilise the power dynamics that privilege the perspectives of users and designers of technologies of violence and allow a better understanding of the nature of terrorism and the ways in which commonly accepted forms of state violence might themselves be forms of terrorism.

## References

1. Latour B (2005) *Reassembling the social: an introduction to actor-network theory*. Oxford University Press, Oxford
2. Amoore L (2009) Algorithmic war: everyday geographies of the war on terror. *Antipode* 41(1):49–69
3. Atherton KD (2020) Trump inherited the drone war but ditched accountability: only a single formal check remains on U.S. killings worldwide. *Foreign Policy*, 22 May 2020. Available at <https://foreignpolicy.com/2020/05/22/obama-drones-trump-killings-count/>
4. Bacchini F, Lorusso L (2019) Race, again: how face recognition technology reinforces racial discrimination. *J Inf Commun Ethics Soc* 17(3):321–335
5. Benjamin R (2019) *Race after technology*. Polity Press, Cambridge, UK
6. Brettschneider C (2007) The rights of the guilty: punishment and political legitimacy. *Polit Theory* 35(2):175–199
7. Center for Civilians in Combat and Columbia Law School Human Rights Clinic (2012) *The civilian impact of drones: unexamined costs, unanswered questions*. Columbia University, New York
8. Coeckelbergh M (2013) Drones, information technology, and distance: mapping the moral epistemology of remote fighting. *Ethics Inf Technol* 15:87–98
9. Cooper H, Schmidt E (2021) Video showed at least 1 child near site minutes before drone strike in Kabul. *The New York Times*, November 3, 2021. <https://www.nytimes.com/2021/11/03/us/politics/drone-strike-kabul-child.html>
10. Cooper H, Schmitt E (2021) U.S. airstrikes in Syria target Iran-backed militias that rocketed American troops in Iraq. *The New York Times*, 25 Feb 2021. <https://www.nytimes.com/2021/02/25/us/politics/biden-syria-airstrike-iran.html>

11. Crawford N (2020) Afghanistan's rising civilian death toll due to airstrikes, 2017-2020. The costs of war project, Brown University, 20 Dec 2020. [https://watson.brown.edu/costsofwar/files/cow/imce/papers/2020/Rising%20Civilian%20Death%20Toll%20in%20Afghanistan\\_Costs%20of%20War\\_Dec%202020.pdf](https://watson.brown.edu/costsofwar/files/cow/imce/papers/2020/Rising%20Civilian%20Death%20Toll%20in%20Afghanistan_Costs%20of%20War_Dec%202020.pdf)
12. Cummings ML (2012) Automation bias in intelligent time critical decision support systems. In: AIAA 1st intelligent systems technical conference, 19 June 2012. <https://doi.org/10.2514/6.2004-6313>
13. Cunneen C (2020) 'The torment of our powerlessness': police violence against aboriginal people in Australia. Harvard International Review, 30 Sept. <https://hir.harvard.edu/police-violence-australia-aboriginals/>
14. Dobos N (2020) Ethics, security, and the war-machine: the true cost of the military. Oxford University Press, Oxford
15. Edwards HL, Esposito M (2019) Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proc Nat Acad Sci* 116(34):16793–16798
16. Gartenstein-Ross D, Clarke CP, Shear M (2020) Terrorists and technological innovation. *Lawfare*, 2 Feb 2020. <https://www.lawfareblog.com/terrorists-and-technological-innovation>
17. Henschke A, Legrand T (2017) Counterterrorism policy in liberal-democratic societies: locating the ethical limits of national security. *Aust J Int Aff* 17(5):544–561
18. Hossain MS (2015) Social media and terrorism: threats and challenges to the modern era. *S Asian Surv* 22(2):136–155
19. International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU Law School (2012) Living under drones: death, injury, and trauma to civilians from US drone practices in Pakistan. <https://www-cdn.law.stanford.edu/wp-content/uploads/2015/07/Stanford-NYU-Living-Under-Drones.pdf>
20. Jackson R (2005) Writing the war on terrorism: language, politics, and counterterrorism. Manchester University Press, Manchester, UK
21. Kaag J, Kreps S (2014) Drone warfare. Polity Press, Cambridge, UK
22. Kershner S (2013) Autonomous weapons pose no moral problems. In: Strawser BJ (ed) Killing by remote control: the ethics of an unmanned military. Oxford University Press, New York, pp 229–246
23. Killmister S (2008) Remote weaponry: the ethical implications. *J Appl Phil* 25(2):121–133
24. Levin S (2020) 'It's not about bad apples': how US police reforms have failed to stop brutality and violence. *The Guardian*, 16 June 2020. <https://www.theguardian.com/us-news/2020/jun/16/its-not-about-bad-apples-how-us-police-reforms-have-failed-to-stop-brutality-and-violence>
25. Liao S, Huebner B (2020) Oppressive things. *Phil Phenomenological Res* (online first). <https://0-onlinelibrary-wiley.com.libus.csd.mu.edu/doi/full/10.1111/phpr.12701#reference>
26. Linders A, Kansal SP, Sharpe K, Oakley S (2020) The promises and perils of technological solutions to the troubles with capital punishment. *Humanity Soc*. <https://journals.sagepub.com/doi/abs/10.1177/0160597620932892?journalCode=hasa1-30>
27. Lucas G (2013) Engineering, ethics, and industry: the moral challenges of lethal autonomy. In: Strawser BJ (ed) Killing by remote control: the ethics of an unmanned military. Oxford University Press, New York, pp 211–228
28. MacFarquhar N (2020) As domestic terrorists outpace jihadists, new U.S. law is debated. *New York Times*, 25 Feb 2020. <https://www.nytimes.com/2020/02/25/us/domestic-terrorism-laws.html>
29. Maguire ER (2015) New directions in protest policing. *Saint Louis Univ Public Law Rev* 35(1):67–108
30. Marx L (1997) Technology: the emergence of a hazardous concept. *Soc Res* 64(3):965–988
31. McReynolds P (2005) Terrorism as a technological concept: how low versus high technology defines terrorism and dictates our responses. In: Shanahan T (ed) *Philosophy 9/11: thinking about the war on terrorism*. Open Court, Peru, Illinois, pp 69–93
32. Michaelson C (2012) The triviality of terrorism. *Aust J Int Aff* 66(4):431–449

33. Perry AM, Harshbarger D, Romber C, Thymianos K (2020) To add value to black communities, we must defund the police and prison systems. Brookings.edu, 11 June 2020. <https://www.brookings.edu/blog/how-we-rise/2020/06/11/to-add-value-to-black-communities-we-must-defund-the-police-and-prison-systems/>
34. Primoratz I (2013) *Terrorism: a philosophical investigation*. Polity Press, Cambridge, UK
35. Primoratz I (2018) *Terrorism*. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy*, Winter 2018 edn. <https://plato.stanford.edu/archives/win2018/entries/terrorism/>
36. Qaurooni D, Ekbia H (2017) The “enhanced” warrior: drone warfare and the problematics of separation. *Phenomenology Cogn Sci* 16(1):53–73
37. Regan MC (2021) Citizens, suspects, and enemies: examining police militarization. *Texas National Security Review*, Winter 2020/2021. <https://tnsr.org/2020/12/citizens-suspects-and-enemies-examining-police-militarization/>
38. Rejali D (2003) Modern torture as civic marker: solving a global anxiety with a new political technology. *J Hum Rights* 2(2):153–171
39. Rhodes L (2007) Supermax as a technology of punishment. *Soc Res* 74(2):547–566
40. Rodin D (2004) Terrorism without intention. *Ethics* 114(4):752–771
41. Savage C, Schmitt E (2021) Biden secretly limits counterterrorism drone strikes away from war zones. *The New York Times*, 3 Mar 2021. <https://www.nytimes.com/2021/03/03/us/politics/biden-drones.html>
42. Schwarz E (2018) Technology and moral vacuums in just war theorizing. *J Int Polit Theor* 14(3):280–298
43. Schwarz O (2020) After the protests, lingering trauma: the scars of ‘non-lethal’ weapons. *The Guardian*, 12 Aug 2020. <https://www.theguardian.com/world/2020/aug/12/george-floyd-protests-lingering-trauma-non-lethal-weapons-scars>
44. Sharkey N (2010) Saying ‘No!’ to lethal autonomous targeting. *J Mil Ethics* 9(4):369–383
45. Sierra-Arévalo M (2019) Technological innovation and police officers’ understanding and use of force. *Law Soc Rev* 53(2):420–451
46. Skerker M, Purves D, Jenkins R (2020) Autonomous weapons systems and the moral equality of combatants. *Ethics Inf Technol* 22(3):197–209
47. Skolnik J (2021) Pentagon watchdog finds no evidence of criminal negligence in “regrettable” Kabul drone strike. *MSN*, Salon, November 5, 2021. <https://www.msn.com/en-us/news/world/pentagon-watchdog-finds-no-evidence-of-criminal-negligence-in-regrettable-kabul-drone-strike/ar-AAQI9Nk?ocid=uxbndlbing>
48. Sparrow R (2007) Killer robots. *J Appl Philos* 24(1):63–77
49. Sparrow R (2013) War without virtue? In: Strawser BJ (ed) *Killing by remote control: the ethics of an unmanned military*. Oxford University Press, New York, pp 84–105
50. Steinhoff U (2013) Killing them safely: extreme asymmetry and its discontents. In: Strawser BJ (ed) *Killing by remote control: the ethics of an unmanned military*. Oxford University Press, New York, pp 179–210
51. Strawser BJ (ed) (2013a) *Killing by remote control: the ethics of an unmanned military*. Oxford University Press, New York
52. Strawser BJ (2013b) Introduction: the moral landscape of remote weapons. In: Strawser BJ (ed) *Killing by remote control: the ethics of an unmanned military*. Oxford University Press, New York, pp 3–24
53. Swartzter S (2019) Race, ideology, and the communicative theory of punishment. *Philosophers’ Imprint* 19(53):1–22, 11
54. UN News (2019) New technologies, artificial intelligence aid fight against global terrorism. <https://news.un.org/en/story/2019/09/1045562>
55. van der Linden H (2016) Arguments against drone warfare with a focus on the immorality of remote control killing and “Deadly Surveillance”. *Radical Philos Rev* 19(2):331–358
56. Verbeek P-P (2011) *Moralizing technology: understanding and designing the morality of things*. University of Chicago Press, Chicago
57. Walzer M (2001) Excusing terror. *The American prospect*, 5 Nov. <https://prospect.org/features/excusing-terror/>

58. Williams JP (2021) The U.S. capitol riots and the double standard of protest policing. US News, 12 Jan. <https://www.usnews.com/news/national-news/articles/2021-01-12/the-us-capitol-riots-and-the-double-standard-of-protest-policing>
59. Wolfendale J (2016) The narrative of terrorism as an existential threat. In: Richard J (ed) The Routledge handbook of critical terrorism studies. Routledge, Abingdon, UK, pp 114–124
60. Woodside SN (2013) Unintentional terrorism? An objection to David Rodin's 'terrorism without intention'. J Mil Ethics 12(3):252–262

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# On the Moral Significance of Narrative, Imagery, and Social Signalling in Counterterrorism Targeted Killing Operations



Michael Robillard

## 1 Introduction

Some philosophers have argued that the use of drones and related UAV technologies in warfare is *in principle* morally problematic.<sup>1</sup> Often, these accounts make the claim that something about the absence of *sufficient risk* to the pilot makes the employment of drones somehow unfair, indecent, or unvirtuous in some respect.<sup>2</sup> Other philosophers, however, have argued that there is in fact no *in principle* reason that speaks against the use of such technologies in warfare, and that under certain circumstances, it might be not just permissible but indeed *obligatory* to employ such technologies.<sup>3</sup> Lastly, there have been many and various *contingent* arguments for and against the use of drones ranging from moral concerns about proliferation, to moral hazard and overuse, to issues of government transparency, to PTS and moral injury experienced by pilots, to dangers of inciting eventual blowback.<sup>4</sup> By now, many of these in-principle as well as contingent ethical arguments surrounding the

---

M. Robillard (✉)

University of Notre Dame, Notre Dame, IN, USA

e-mail: [michaelrobillard@protonmail.com](mailto:michaelrobillard@protonmail.com)

<sup>1</sup> Since it has become such a pervasive part of our contemporary lexicon, I will use the canopy term ‘drones’ to refer to the general set of U.S. un-manned aerial targeted killing platforms (i.e. Predators, Reapers, etc.). Also, for the sake of this chapter, I will refrain from speaking about so-called ‘fully-autonomous’ weapons since they entail their own set of weighty metaphysical and moral complications orthogonal to this particular debate.

<sup>2</sup> For arguments of this sort see Daniel Brunstetter and Megan Braun, “The Implications of Drones on the Just War Tradition,” *Ethics and International Affairs*, Volume 25, Issue 3, Fall 2011, pp. 337–358.

<sup>3</sup> For arguments in this sort, see Bradley Strawser “Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles,” *Journal of Military Ethics* 9, no. 4 (December 2010): 342–368.

<sup>4</sup> For treatment of contingent arguments related to drone proliferation see Robert Buchanan and Robert O. Keohane, “Toward a Drone Accountability Regime” *Ethics and International Affairs*, Spring 2015, pp. 15–37.

employment of drones constitute well-trodden and familiar ground for many just war theorists.

Despite the vast and rich amount of philosophical literature surrounding the ethics of drones that has been generated over the past decade and a half or so, I believe that there is something still left to be said with respect to the morality of drone use for *counterterrorism operations* specifically. In particular, for the aims of this chapter, I wish to build off of two arguments already developed in this space by Rebecca Johnson and Tom Simpson respectively. If we take the *likelihood of success* criteria to be a necessary feature of fighting a just war, *and* if we regard counterinsurgency and counterterrorism operations to admit of a *fundamentally different* set of success conditions when compared to conventional-style warfare, then questions concerning the permissibility or impermissibility of drone use will (at least in part) be determined by the degree to which their use serves or detracts from the overall realization of such success conditions.

Since success in counterterrorism operations fundamentally involves ‘winning hearts and minds’, building enduring, on-the-ground trust relationships, and maintaining narrative dominance, all while denying one’s adversaries the ability to do the same, then our overall ethical appraisal of the use of drones in such operations must take such success conditions into account in a way *that wouldn’t otherwise be taken into account* in a conventional war context. Part-and-parcel of this ethical appraisal then is proper regard for the moral significance that the non-kinetic features of *narrative*, *imagery*, and *social signalling* play with respect to remote targeted killing operations. Lastly, these factors are made even more ethically and strategically important when considering the fast rise of social media and smartphone technology on the battlefield.

The structure of this chapter will proceed as follows. In Sect. 2, I will explore some of the generally accepted doctrinal wisdom with respect to contemporary U.S. counterterrorism and counterinsurgency operations. In Sect. 3, I will unpack some of the relevant features of Johnson and Simpson’s arguments and explore the two different and nuanced perspectives they bring to the discussion regarding ethical features of drone use and trust building in counterterrorism operations. In Sect. 4, I will make the case for the increasing moral (and strategic) weight of *narrative*, *imagery*, and *social signalling effects* with respect to *in bello* proportionality. And in Sect. 5, I will explore how these normatively-laden non-kinetic factors fit into our moral and strategic thinking about drone use in contemporary counterterrorism operations specifically.

## 2 Irregular Warfare

In its simplest form, irregular warfare, both offensively and defensively, involves, to quote Mao Zedong, ‘winning hearts and minds’ [12]. For insurgents, this goal is usually accomplished by means of using asymmetric or guerrilla warfare tactics (e.g. I.E.D.s, snipers, mortars, etc.) against the conventional military and police forces

of an existing regime or governance in order to weaken or discredit that governance's overall effectiveness and trustworthiness in the eyes of the general populace. Other times this goal is accomplished by using terror tactics against innocent non-combatants of a given populace in order to intimidate that populace into compliance and/or to discredit the existing regime's legitimacy and capacity to protect its own people.<sup>5</sup> Conversely, counterinsurgency and counterterrorism operations seek to establish, maintain, and improve well-ordered, legitimate, and trustworthy governance over a given populace while simultaneously disrupting, destroying, and denying sanctuary to such insurgent and terrorist threats.

From a Maoist 'protracted warfare' paradigm, terrorist and insurgent groups aim to move from a 'guerrilla/insurgency' phase to a 'positional' phase involving increased control and influence over physical terrain, infrastructure, local populace, conventional forces, and state apparatus, to a third and final phase involving conventional power projection outward and beyond state borders. The goal of counterinsurgency operations is therefore the reverse; to deny enemy power projection, to gain dominance over positional assets, and to increasingly put insurgent adversaries on the defensive back into guerrilla mode, until they lack the physical means and/or will to continue to fight or organize [3].

Importantly, the Clausewitzian 'center of gravity', so to speak, in this perpetual struggle between insurgents and counterinsurgents is to be fundamentally located in the trust and sympathies of *the local populace* and much less so in the conventional space of men and material. Indeed, Harry Summers classic, *On Strategy: A Critical Analysis on the Vietnam War* highlights this very point in his analysis of the U.S. war in Vietnam. Despite the United States winning every single conventional *kinetic* battle during Vietnam, it nonetheless lost the war. Hence, contra traditional Clausewitzian thinking, the mere aggregation of kinetic victories in battle did not equate to the winning of the war in total. Rather, when it came America's lost war in Vietnam, winning the 'hearts and minds' of the local populace is what fundamentally mattered in the end [9].

The last two decades-worth of American warfighting efforts in the Middle East and beyond have attempted to make good use of these hard-fought lessons from Vietnam. Contemporary doctrinal thinking on irregular warfare and counter-insurgency reflects this shift in conceptualizing conflict, war, and conditions of success and failure for such nuanced contexts. Central to this conceptual space of irregular warfare is the centrality of *narrative* in effective counter-terrorism operations. The recently updated edition of JP 3-24, *Counterinsurgency*, acknowledges this very fact stating the following,

In the context of insurgency, the narrative is a tool to shape how the population perceives circumstances and events. The narrative is used to link conditions-based grievances to the nature or behaviour of the incumbent regime and articulate an alternative political vision

---

<sup>5</sup> I wish to avoid getting into an in-the-weeds debate here about counterinsurgency versus counterterrorism. For the sake of this paper then, I will use these terms largely interchangeably with 'terrorism' being a particular violent method used by insurgents, directed at innocent civilians (as opposed to police and soldiers) to achieve their political ends.



that will address those grievances. It provides an explanation and justification of how insurgents will align ends, ways, and means to achieve their political objectives and frames how insurgent and counterinsurgent actions are interpreted.

*JP 3-24* continues,

The likelihood of insurgent success is based in large part on assessments of insurgent political and military strength. The uncertainty inherent in insurgency, coupled with the competition between insurgent propaganda and counterinsurgent information-related activities, often generates wild rumors and distorted perceptions of particular incidents. Populations can often only assess that strength in their immediate vicinity, generating wildly different perceptions of the broader national environment in different parts of the operational area. [5]

While far from being exhaustive, this unique set of strategic, operational, and tactical considerations constitutes the general conceptual space that contemporary military ethicists and just war theorists must focus on if they are to speak meaningfully and productively with regard to prescriptive moral guidance for counterinsurgency and counterterrorism operations. Such prescriptions must therefore take into account the moral significance of not just typical kinetic harm trade-offs but also the moral significance of *non-kinetic* factors such as social signalling, imagery, and narrative.

### 3 Broad Counterterrorism Ethics Considerations

In her paper, “The Wizard of Oz goes to War: Unmanned Systems in Counterinsurgency,” Rebecca Johnson notes the various moral goods that could be gained by prudent and responsible use of drone assets in counterterrorism and counterinsurgency contexts. Some of these moral goods include; (1) the potential to more effectively achieve our just ends (2) the mitigation of soldier risk, and (3) the ability to make better calculations regarding necessity, proportionality, and discrimination in the absence of immediate battlefield duress. Lastly, a final and novel moral good that Johnson points out, one often overlooked, is the surplus time and energy that effective drone use could free up for soldiers on the ground. In other words, if drones could be used effectively to achieve a baseline of security and protection, then such protection could then free up valuable time, energy, and effort for soldiers and commanders on the ground to shift to governance and institutional-building operations [4]. She writes, “so as long as humans guiding and prosecuting the war remain committed to the principles of combatancy, distinction, and non-combatant immunity- and at present we have no indications to the contrary-unmanned systems improve civilians’ and military personnel’s ability to prosecute counterinsurgency effectively and morally.” [4]

While I am in general agreement with Johnson’s claim, I have begun to grow increasingly skeptical about how we should conceive of satisfying conditions of non-combatant immunity in light of downstream social effects due to imagery and narrative. Johnson gestures at something similar when she writes, “It is not enough to minimize collateral damage in a literal, body count, sense; US forces have the

additional responsibility of minimizing the effect their presence has on the fabric of the civilian population to protect the population's ability to return to a state of peace following the war." [4]

Hence, if it turns out that our use of drones in counterinsurgency missions succeed in minimizing collateral damage in the 'body count' sense, as Johnson puts it, but in so doing creates social events that degrade the social tapestry of the local population, then we must reconsider *both* our conceptions of collateral damage as well as mission success.

In his paper, "Robots, Trust, and War," Tom Simpson gestures at similar moral and strategic tensions concerning the deleterious effects that the employment of *fully* autonomous weapons might have for establishing trustworthy relationships in counterterrorism and counterinsurgency contexts. He writes,

The will of the population constitutes, in the terms of modern NATO doctrine, the strategic centre of gravity. Winning their trust wins you the war. The host population must trust that you will act in a way that takes their interests fully into account, and furthermore, must trust that you will defeat the insurgents. So the trust involved is dynamically interactive. And host populations through history have certainly exhibited reactive sentiments of anger when they have felt betrayed or badly used by expeditionary forces, and correspondingly, have shown gratitude when expeditionary forces have prevailed in a way that they have welcomed. For both these reasons, it must be the host population's *normative* trust that must be won, and not (solely) predictive." [8]

By 'center of gravity', Simpson is referring here to the Clausewitzian notion of an enemy's main source of power that provides moral and physical strength and the will to act [1]. While Simpson's argument is technically against the use of *fully* autonomous weapons, I believe that the general spirit of his argument still applies to the potentially deleterious social and narrative effects that present-day *semi-autonomous* drone use might have on our long-term ability to create lasting and enduring trust-relationships as part of our ongoing counterterrorism efforts.

While I am in general agreement with the overall thrust and focus of Johnson and Simpson's accounts, I believe that the sharp rise in social media and overall informational connectivity in the world since the time of their respective publications has made it the case that the debate regarding the ethics of drones in counterterrorism operations must be updated and re-contextualized with a much more significant weight being granted to narrative, social signalling, and imagistic factors. I will now explain what I mean by this.

#### **4 The Moral Significance of Narrative, Social Signalling, and Imagery**

The idea that the signalling or communicative effects of a self-defensive or other defensive act can have moral or justificatory force is nothing new for just war theorists. Indeed, it has been a long-held belief by many theorists that the 'future deterrence value' of a violent act or threat of a violent act, on both the individual and nation-state

level can be morally justified under certain circumstances. On an individual level, in the absence of an effective police force, it is arguably morally justified for the isolated Afghan goat-herder, for instance, to respond with disproportionately lethal force *at the moment* (say, by shooting a thief attempting to steal one of his goats) if that act will foreseeably deter a much *greater future harm* from foreseeably occurring later on (say, inviting a gang of thieves to rape and pillage his home a week later if he does nothing).

Theorists also recognize the justificatory force that future deterrence value adds at the nation-state level as well. Indeed, the idea that the *threat of violence* can be good insofar as it serves to deter a much greater future harm from actualizing down the line is presumably why we think nation-states can permissibly build up armies in the first place and why we think nations can sometimes justifiably launch pre-emptive attacks. These individual and collective cases therefore give credence to the moral significance of social signalling effects built into kinetic acts of harming. However, future deterrence in particular seems often to be the very limited sense in which many just war theorists regard the signalling value of violence to morally matter. Indeed, there seems to be other morally important ways; ways beyond just future deterrence, in which the signalling value of a violent act in war can morally matter. As LTC Bob Underwood notes,

Killing in war eliminates threats but also plays a part in influencing the decisions of other persons beyond those we might kill. This suggests that killing in war has a communicative function, and that the message is an important consideration that can feature in the balance of reasons to kill some but not others in war. This is true provided combatants can permissibly kill some as means to communicate to others. I argue that just combatants, those that fight for just aims, can permissibly kill to communicate and that unjust combatants cannot. This is a new reason to revise our intuition that combatants on both sides hold equal rights to kill, the so-called moral equality of combatants (MEC). [11]

This is an important point here that Underwood brings up. However, I believe that it is even more important when we consider such claims against the backdrop of a counter-insurgency/counterterrorism paradigm; one saturated with the morally salient features of trust-building and winning hearts and minds that Johnson and Simpson both high-light. That being said, we could go even further and say that not only does *killing itself* have communicative or signalling value that morally matters with respect to *in bello* proportionality, but that *the way, look, and social presentation* in which the killing is done has similar or greater communicative or signalling value as well.<sup>6</sup>

For instance, consider the social signalling effects of the 2013 U.S. drone strike against suspected Al Qaeda operatives in Yemen that turned out to be a wedding procession versus the successful ‘boots on the ground’ Special Forces raid to kill Abu Bakr Baghdadi, head of ISIS, in 2019 [10]. Epistemic uncertainties and kinetic trade-offs aside, the first case arguably would have still given terrorist organizations

---

<sup>6</sup> By *in bello* proportionality I mean the trade-off between predicted goods and harms for a particular act *in battle*. We can contrast this with *ad bellum* proportionality, the predicted trade-off of goods and harms having to do with a nation-state choosing to go to war at all.

narrative and imagistic fodder even if the target turned out to be actual Al Qaeda operatives whereas the second case would have provided much less narrative and imagistic fodder even if collateral damage was taken.

These considerations arguably change our typical thinking when it comes to the ethics of drone use, particularly within counterinsurgency and counterterrorism contexts, especially when considered in light of (1) the mid and long-term communicative and narrative effects on the local populace, and (2) the potential for narrative and imagery surrounding such eliminative killing acts to be leveraged by terrorist groups for their own propagandistic purposes.<sup>7</sup>

Consider then the following two mission options.

*Option 1:* I use my drone intelligence capabilities to locate an unjust, fully liable high-value-target (HVT) who is located far from any noncombatants. I then shoot the HVT with my drone.

*Option 2:* I use my drone intelligence capabilities to locate an unjust, fully liable HVT who is located far from noncombatants. I then deploy my fully willing SEAL team to enter into a space of increased (but still manageable risk) and then shoot the HVT.

Looked at strictly in terms of eliminative harming trade-offs, ‘body count’ conceptualizations of collateral damage and in the absence of consideration for narrative and signalling effects, it seems questionably permissible for a commander to choose Option 2 over Option 1. If one has the means to eliminatively kill a liable target, with no likelihood of (physical) collateral damage, and with overall less risk to soldiers, then what additional moral reason would one have to needlessly risk soldier lives, even if it turned out the soldiers willingly volunteered to take such risk?

However, once we begin taking into account the moral importance of narrative, social signaling, and imagery in the short, mid, and long-term, our moral calculus arguably changes especially for counterinsurgency/counterterrorism contexts. If, in the short-term, a drone-strike serves to successfully take out five fully liable high value terrorist targets with no collateral damage whatsoever, but the narratives, local rumors, and imagery from the event can be more quickly and effectively reappropriated by terrorist actors to convince the local populace *that there was* in fact major collateral damage to innocents, then it is hard to count the drone strike as prudentially or strategically sound. At best, it seems like a short-term tactical win taking away from mid to long-term operational and strategic success. And insofar as the *in bello* action diminishes the overall likelihood of success, the act seems questionably moral as well. Were it the case that we knew that part-and-parcel with a kinetic drone strike we would also be providing terrorist actors narrative and imagistic fodder for future propaganda use deleterious to our overall goal of winning hearts and minds, then, all other things being equal, a combination of ethics and prudence might begin to nudge us away from such drone options. We might instead find it morally and strategically

---

<sup>7</sup> Arguably, another theoretical option on the table would be a strategy to completely normalize drone killings so that they no longer have such negative downstream social effects. For instance, the first-time guns were used in war their use arguably had a serious narrative and messaging effect associated with them that, after a period of normalization, went away. I am quite wary of such normalization however.

preferable to employ soldiers on the ground at a heightened chance of risk for the sake of the positive narrative effects on the populace such an act would yield or at least for the negative narrative effects on the populace such an act would deny to terrorist adversaries.

At first glance, the argument I am making here sounds identical to those philosophers arguing for the in-principle impermissibility of drones based upon the absence of soldier risk.<sup>8</sup> This however is not the kind of argument I am advancing here. Indeed, I believe that such accounts are incorrect and that these types of arguments for the *prima facie* impermissibility of drones because they are ‘riskless’ actually get the order of moral justification completely reversed. Indeed, soldiers derive their justification for fighting and killing in the first place from the justness of their nation’s cause, not from bootstrapping *ad bellum* moral justification *ex nihilo* from out of physical risk on the battlefield. It can’t be the case that soldier physical risk *itself* is necessary in order to fight a just war and/or that such risk is the source of moral justification for *in bello* harming. Otherwise, it would logically follow that whichever political project (ISIS, Nazi Germany, Soviet Russia, etc.) subjected their troops to *more* physical risk, intentionally or just by accident, would necessarily have the claim to the moral high-ground. Such a position would also entail the radical implication that since the advent of the shield and the bow and arrow, wars have been getting progressively *more and more unjust* in lock-step with each technological advancement in weaponry and defensive capacities that increased lethality while mitigating soldier risk. That can’t be the case.

That being said, there is still something important that soldier-risk proponents are getting at and I believe it is the following. Physical risk *qua* physical risk doesn’t itself provide any additional moral justification for harming and war. However, the *communicative* and *social signalling* effects constitutive of demonstrations of physical risk in battle *do* in fact have moral weight. This is a subtle but important distinction since it is the set of positive downstream goods generated by the signalling act of risk and not *the act of risk itself* which is the source of moral justification force certain kinds of harming acts. The weight of such signalling acts is particularly morally significant if we consider the potentially beneficial or deleterious second and third order effects such signalling acts will have on our project of ‘winning hearts and minds’ and, conversely, on terrorist groups’ project of doing the same. If we begin scooping those downstream second and third order social effects into our *in bello* proportionality calculus, then the case for preferring Option 2 over Option 1 begins to find greater moral and prudential appeal. However, it is admittedly tough if not impossible to count these social downstream effects as part of our *in bello* proportionality calculus beyond a certain predictive window. In particular, given that these

---

<sup>8</sup> For arguments of this sort see works by Paul Kahn, Christian Enemark and Anders Henriksen & Jens Ringsmose: Kahn, Paul W. "The paradox of riskless warfare." *Philosophy and Public Policy Quarterly* 22.3 (2002): 2–7; Enemark, Christian. *Armed drones and the ethics of war: military virtue in a post-heroic age*. Routledge, 2013; Henriksen Anders & Ringsmose, Jens (2015) Drone warfare and morality in riskless war, *Global Affairs*, 1:3, 285–291, <https://doi.org/10.1080/23340460.2015.1080042>.

second and third order social effects are really hard to predict with certainty, to quantify, and to demonstrate causally, it is arguably much more difficult to count these complicating factors in our proportionality calculations than it is for us to factor brute kinetic effects and trade-offs.

That being said, it is not as if predicting collective human social behavior is completely and totally opaque to us at all times. Accidentally drone striking a mosque or wedding amidst a smart-phone laden public, for instance, will generate certain predictable downstream social effects that are harmful and disruptive to a populace and a mission in a way that goes well beyond the immediate blast radius of the chosen munition. Running a clandestine operation under the cover of darkness will yield another. And while it is arguably incorrect for us to predict every future downstream social harm arising from a singular battle-field act, it is likewise equally incorrect for practitioners and ethicists to treat each kinetic action as somehow occurring in a social vacuum with *no* regard for downstream social and narrative effects whatsoever. My point here is simply that narrative effects need to be weighed more heavily both ethically and strategically given the twenty-first century informational space in which we now find ourselves.

## 5 Application to Counterterrorism Drone Operations

As to the actual extent to which U.S. drone strikes versus precision boots-on-the-ground missions have served to ‘win hearts and minds’ while preventing terrorist groups from doing the same over the past ten years, I do not know. Answering such a question thoroughly is largely an empirical matter and one in which I leave for the anthropologists, political scientists, and people with a higher security clearance than mine to sort out. That being said, I believe there is at least some preliminary empirical evidence to suggest that we ought to start paying more attention to the downstream social signalling effects that our present drone operations might be having with respect to winning hearts and minds and with respect to denying terrorist actors the ability to do the same.

In a recent paper involving an analysis of eighty-seven face-to-face interviews with Afghan civilians affected negatively by US combat operations, Janina Dill explores how civilians directly affected by collateral damage perceive the overall justness of such actions.<sup>9</sup> Despite explaining to these interviewees nuanced international law of armed conflict standards, concepts of proportionality and necessity, and the standard risk mitigation measures that went into such operations, Dill’s report claims that seventy out of eighty-seven of the interviewees, roughly 80 percent, still believed that the coalition had *deliberately* set out to harm them [2]. Several strong and visceral testimonies emphasize this distrust,

---

<sup>9</sup> It is important to note here that not all of Dill’s interviewees suffered collateral harm from drone strikes exclusively. Rather, testimonies primarily involved harms directly due to air strikes, cross-fire incidents, direct shootings, and indirect artillery fire.

We have been told that American technology is so advanced that they can see a needle from the air. Why then don't they distinguish civilians from Taliban? Americans are able to recognize black and white chickens from the air, how come they can't recognize women and children?

Americans are against Muslims. For them, Taliban and civilians are the same.

They are here to kill us and destroy our houses.

They think we are animals. [2]

The testimonial data from Dill's interviews should matter greatly in our thinking about the ethics of drones (and other harming) in counterterrorism operations. What seems to be at issue here is the sharp divergence from official just war wisdom and IHL/LOAC standards concerning such things as proportionality, necessity, distinction, collateral harming, etc. and the indigenous populace's *subjective perception* of such notions. Insofar as the indigenous populace is the centre of gravity for successful counterterrorism and counterinsurgency operations, then their subjective perceptions of the fairness of certain battlefield harms to themselves, their neighbours, or their surrounding community must be taken into account *as we find them*. Otherwise, turning a deaf ear to such perceptions and testimonies results not only in a missed opportunity for building local trust, but also presents a corresponding opportunity for insurgents and terrorist groups to leverage such sentiments towards their own unjust ends.

Several recent articles and reports have suggested that terrorist and insurgent groups are making just such informational and propagandistic moves, both in theatre, in the US, and in the greater Muslim world. For instance, with respect to 'in theatre' propagandistic leveraging, one 2013 *Guardian* article suggested that the US drone program, despite its tactical successes, was fundamentally sowing the seeds of strategic and international failure. In, this article four former US air force members with more than 20 years of combined experience operating drone weaponry systems spoke harshly against the strategic short-sightedness of such a program and the long-term propagandistic and radicalization tool that such a program was handing over to terrorist actors. Several of them went on to make the exceptionally bold claim that the killing of innocent civilians in drone strikes has acted as one of the most, "devastating driving forces for terrorism and destabilization around the world." [7]

A similar trend can be seen with respect to propagandistic efforts directed at Muslims in western-speaking countries and beyond. In "The Portrayal of Drones in Terrorist Propaganda: A Discourse Analysis of Al Qaeda in the Arabian Peninsula's *Inspire*", Jan Andre Lee Ludvigsen outlines some of the elements of Al Qaeda's strategic media messaging campaign surrounding US drone strikes. Broadly speaking, Ludvigsen's report finds that the magazine frequently portrays the US drone campaign as, (1) an ultimately failing policy that has mainly resulted in civilian deaths (2) a tool to oppress Muslims, and (3) a fundamentally cowardly, dishonourable, and inhumane way of fighting. Such appeals to soldierly honor (or the lack thereof) are encapsulated in the following lines from the 10th Edition of *Inspire*:



Let us not forget that America adopted the drone program because this has no costs in terms of (American) lives lost. Successive American administrations have realized that the American soldier is too much of a coward to prove his mettle in wars. [6]

Ludvigsen concludes his article by arguing against the possibility of drawing a *precise* causal or predictive connection between US drone strikes and eventual ‘blowback’ due to radicalization and propagandistic efforts. However, he argues that such increasingly sophisticated propagandistic moves by terrorist and insurgent groups gives US strategists some reason to rethink present and future-facing counterterrorism drone operations [6].

As to the overall effectiveness of such propagandistic efforts to recruit, radicalize, and/or inspire eventual blowback, the jury still seems out. Indeed, much more empirical work, analysis, and predictive modelling will need to be brought to bear on such problems. That being said, the purpose of this paper isn’t to decisively settle these issues. Rather my aim has mainly been to make clearer what the actual moral and prudential reasons are on the moral ledger and how they trade off against one another within a 2020s counterterrorism paradigm. The rather modest thesis of this chapter then is simply that the non-kinetic factors of narrative, social-signalling, and imagery connected to otherwise *kinetic* drone strikes needs to be factored more heavily into our overall moral thinking both with respect to *in bello* proportionality as well as with respect to our mid to long-term strategic thinking. Bearing in mind the amplification of these non-kinetic factors due to the increased speed, reach, and connectivity of the internet and social media and bearing in mind the increased potential for propagandistic leveraging of these non-kinetic factors by insurgent and terrorist actors, ethical thinking about drone use in counterterrorism operations requires reconceptualization in these more fine-grained moral and prudential terms.

## 6 Conclusion

As the author William Morris once said, “nothing useless can be truly beautiful.” I believe something similar can be said regarding ethics, insofar as nothing useless can be truly moral. Indeed, for morality to matter, it must, at some point, find traction with the real-world. In this sense, ethics should inform and mutually reinforce efficacy and efficacy should inform and mutually reinforce ethics. For this to be done successfully, mutual respect and dialogue between theoreticians and practitioners must be accomplished, fostered, and sustained. This is particularly important when it comes to our ongoing ethical and strategic thinking regarding counterterrorism.

When it comes to the ethics of drone use in counterterrorism, we must take the moral significance of non-kinetic factors such as imagery, narrative, and social signalling to weigh more heavily on the moral ledger than we have in the past with respect to our thinking about *in bello* proportionality. While these moral reasons do not necessarily cancel out or override other morally relevant *in bello* factors having to do with such things as mitigating soldier risk, protecting non-combatants, and



ensuring mission success, the increased moral weight of social signalling, imagery, and narrative effects due to the rise of the internet and social media may nonetheless force us to re-evaluate our ethical and strategic thinking about future-facing counterterrorism scenarios heading into the next decade.

## References

1. von Clausewitz K (2009) *On war: the complete*. Wildside Press, Rockville, p 144
2. Dill J (2019) Distinction, necessity, and proportionality: Afghan civilians' attitudes towards Wartime Harm. *Ethics Int Aff* 3(3):315–342. 316–318
3. Guevara C (1963) *Guerrilla war method* available at: <https://www.marxists.org/archive/guevara/1963/misc/guerrilla-war-method.htm>. Accessed 15 July 2020
4. Johnson R (2013) The wizard of Oz goes to war: unmanned systems in counterinsurgency. In: Strawser BJ (ed) *Killing by remote control: the ethics of an unmanned military*, vol 6. Oxford University Press, Oxford, p 20
5. Joint Publication 3-24 (2018) *Counterinsurgency*. Accessed at: [https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3\\_24pa.pdf](https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_24pa.pdf)
6. Ludvigsen JAL (2018) The portrayal of drones in terrorist propaganda: a discourse analysis of Al Qaeda in the Arabian Peninsula's inspire. *Dyn Asymmetric Conflict* 11(1):26–49
7. Pilkington E, MacAskill E (2015) Obama's drone war a 'recruitment tool' by ISIS, Say US air force whistleblowers. *Guardian*. Accessed at: <https://www.theguardian.com/world/2015/nov/18/obama-drone-war-isis-recruitment-tool-air-force-whistleblowers>
8. Simpson T (2011) Robots, trust, and war. *Philos Technol* 24(3): 325–337. 330–331
9. Summers H (1982) *On strategy; a critical analysis of the Vietnam war*. Presidio Press, Novato, pp 1–10
10. Tayler L (2019) A wedding that became a funeral. *Human Rights Watch.org*. Accessed at: <https://www.hrw.org/report/2014/02/19/wedding-became-funeral/us-drone-attack-marriage-procession-yemen>
11. Underwood B (2019) Can soldiers justify killing some as a means to influence the decisions of others? Accessed at: <http://www.bioethics.net/2019/03/oxford-uehiro-prize-in-practical-ethics-question-can-soldiers-justify-killing-some-as-a-means-to-influence-the-decisions-of-others/>
12. Zedong M (1938) *On protracted war*. Accessed at: [https://www.marxists.org/reference/archive/mao/selected-works/volume-2/mswv2\\_09.htm](https://www.marxists.org/reference/archive/mao/selected-works/volume-2/mswv2_09.htm)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Sunlight Glinting on Clouds: Deception and Autonomous Weapons Systems



Amanda Sharkey and Noel Sharkey

*“All warfare is based on deception”* Sun Tzu, 500 BC.

**Abstract** The art of deception has played a significant role in military conflict for centuries and has been discussed extensively. Yet there has been an absence in the literature of any scrutiny of the risks posed by the deception of Autonomous Weapons Systems (AWS). After explaining the nature of AWS, we overview reasons given in their favour and arguments against them. Examples of military deceptive strategies are considered, together with reflections on the nature of deception. The core of the paper is a technical examination of some of the ways that AWS could be deceived and the potential humanitarian consequences. Since AWS have, by definition, an absence of meaningful human control, any deception could remain hidden until too late. We conclude that awareness of the vulnerability of sensing and image processing systems of AWS to deception reinforces and strengthens the case against their development and use.

## 1 Introduction

Sun Tzu’s influential ‘The Art of War’, written in 500 BC and still taught in today’s military academies, stated that ‘all warfare is based on deception’ [1, p. 10]. This is supported by a long history of deception in armed conflict that shows it to be a key to achieving victory. Famous examples include Hannibal’s deception of Flaminius the Roman Consul; luring his troops into ambush [2]. The Confederate army used decoy cannons made from tree trunks [3]. In World War 2, deception about the presence of a (non-existent) superior force waiting outside a harbour led to the Germans sinking their own battleship, the Admiral Graf Spee, in 1939 [4]. In 1944, the US deployed a ‘ghost army’ of 1100 men including artists, architects and set designers to successfully impersonate other Allied Army units, using inflatable tanks, cannons,

---

A. Sharkey (✉) · N. Sharkey  
University of Sheffield, Sheffield, UK  
e-mail: [a.sharkey@sheffield.ac.uk](mailto:a.sharkey@sheffield.ac.uk)

© The Author(s) 2021  
A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,  
Advanced Sciences and Technologies for Security Applications,  
[https://doi.org/10.1007/978-3-030-90221-6\\_3](https://doi.org/10.1007/978-3-030-90221-6_3)

aeroplanes, and both sonic and radio deception. They participated in many operations, impersonating various, larger US units [5]. The D-day invasion of Normandy relied on many deceptions by the Allies about the expected landing location—double agents and other deceptive strategies strengthened the German belief that it would occur at Pas-de-Calais and in Norway [6].

Our question here is, how does deception fit into the ongoing technological transformation of warfare where ever more control of weapons is being ceded to computer systems? In particular, we explore the risks posed by deception to the deployment of Autonomous Weapons Systems. Despite the strong links between deception, deceptive strategies and military operations, there has been little or no discussion about deception and weapons that are entirely controlled by computer algorithms. We begin with an account of Autonomous Weapons Systems (AWS), some of the driving forces in their favour, and the main arguments against them. This is followed by an overview of military uses of deception and deceptive strategies and some reflection on what counts as deception. We then turn to a consideration of how AWS are likely to impact on and be affected by deception in armed conflict and counter-terrorism.

## 2 Autonomous Weapons Systems

Autonomous weapons systems represent a comparatively recent development. The US Department of Defence defines them as weapons that are able, ‘once activated, to select and engage targets without further intervention by a human operator’, [7, updated 2017]. Human Rights Watch [8] defines them as weapons which ‘would identify and fire on targets without meaningful human control’, and the International Committee of the Red Cross defines them as ‘weapons that can independently select and attack targets’ [9].

Since 2013 there have been vigorous attempts at the United Nations Convention for Certain Conventional Weapons (CCW) to create a new international legal instrument to prohibit the development and use of AWS [10, 11]. In 2016, the CCW agreed to establish a Group of Governmental Experts (GGE) to discuss autonomous weapons. Yet superpowers and other major military powers seem to be embarking on their development in the belief that they will create a military edge [12]. China, Russian, Israel and the US wish to use them for force multiplication with very few human controllers operating swarms of weapons in the air, on the land as well as on and under the sea [11].

Major military powers advocate the use of AWS for armed conflict and counter-terrorism for They hold that (i) AWS could complete missions in environments where communication signals could be jammed or disrupted (e.g. [13]), (ii) armed conflict is becoming too fast for human decision making [12], (iii) they could reduce risks for military personnel (e.g. [14]), (iv) they could increase target accuracy and reduce risks for civilians [15].

All four reasons are problematic. In order (i) means that there is no possibility of human oversight to disengage from erroneous targets; (ii) the speed of armed conflict

will simply get faster in a speed race if AWS are used; (iii) they will not reduce the risks to military personnel if the enemy also has them; (iv) while they may increase target accuracy, this does not solve the problems of target selection—the wrong targets could be accurately killed.

More than 140 humanitarian disarmament organisations from 60 countries, clustered under the banner of the Campaign to Stop Killer Robots, are concerned about the harm that such weapons would have on civilian populations (<https://www.stopkillerrobots.org> consulted April 15 2021). Their concerns are expressed in a plethora of arguments that we can break down into 4 major classes: (i) *Non-compliance with International Humanitarian Law (IHL)*: AWS cannot be guaranteed to comply with IHL; (ii) *Immoral delegation of kill decision*: delegating the decision to kill to a machine is immoral (iii) *Global security*: the widespread use of weapons outside of human control would destabilize global security; (iv) *Algorithmic injustice*: algorithmic bias and injustice: the widespread use of algorithms in civil society have shown decision biases against women, ethnic minorities and people of colour and are resulting in many legal challenges. We add to these arguments with an analysis of how susceptible AWS are to deceptive military strategies. Since AWS operate without meaningful human control, they could be subject to deceptions that a human might have detected. They can also, as explained, be deceived by visual manipulations undetectable to the human eye.

### 3 Arguments Against the Use of Autonomous Weapons Systems

International Humanitarian Law (IHL), sometimes termed the Laws of War, is intended to protect civilians. The principles of distinction, proportionality and military necessity are crucial aspects of IHL. Arguments about the inability of AWS to always comply with IHL have been made by Noel Sharkey [11, 16]—see (i) above. The principle of distinction in IHL requires weapons systems to distinguish combatants from non-combatants and other immune actors. Sharkey [16] argues that AWS lack three necessary components for this. First, their sensory and vision systems are not able to reliably discriminate between combatants, non-combatants and other immune actors such as wounded combatants. Second, there is no codified or programmable definition of what constitutes a civilian or non-combatant. And third, AWS lack the necessary situational and battlefield awareness. For instance, a human could draw on their understanding of social situations to recognise insurgents burying their dead in a way that AWS could not.

Sharkey [16] also claims that the principles of proportionality and military necessity are beyond the capabilities of present and near future weapons systems. Some proportionality problems are relatively easy to solve, and some are much harder. The easier proportionality problems involve calculations such as working out the likely collateral damage of different forms of attack and minimising such damage. For

instance, AWS software could choose the munitions to be used near a school so as to minimise the number of children killed. Hard proportionality problems are those which involve decisions about military advantage and military necessity—in the school example, deciding whether the military advantage to be gained would justify the use of any form of attack near a school. Such decisions require, ‘responsible accountable human commanders, who can weigh the options based on experience and situational awareness’ [16]. Suchman [17] has also argued that machines cannot fulfil the requirement of situational awareness. And Sharkey [18] has argued that it is not possible to program or train computational devices to develop the necessary moral competence to make such decisions.

Other writers such as former UN special rapporteur Heyns [19] and Asaro [20] have focused more on the moral argument, (see (ii) above), and argued that irrespective of what AWS and Artificial Intelligence might be able to do in the future, there are important arguments to make about what they should and should not do. For Heyns, AWS *should not* be used to target humans because their use would be an offence against the right to life. Heyns argues that errors would be made and there would be no person to be held accountable. He also claims that the lack of human deliberation would render targeting decisions as arbitrary and against the right to dignity of those targeted and of those in whose name the force was deployed (see also [21]). Asaro [20] argues that AWS *should not* be used even if they were able to meet the requirements of international humanitarian law. For him, IHL and the principles of distinction, proportionality and military necessity, imply a requirement for *human* judgement, and a duty not to delegate the capability to initiate the use of lethal force to unsupervised machines or automated processes.

As well as concerns about the extent to which AWS can conform to IHL, the third set of arguments against them, (iii above), is that they will destabilise global security. Although it is sometimes claimed that AWS could result in more accurate targeting and freedom from human self-preservation concerns, Tamburrini [22] articulates what he terms a ‘wide consequentialist view’ that AWS will threaten global security. He agrees with Sharkey [23] that by reducing the risks of a ‘body bag count’, a major disincentive for war would be removed. Tamburrini [22] also argues that swarms of AWS could weaken traditional nuclear deterrent factors by means of the threat of destructive attacks on strategic nuclear sites that could eliminate an opponent’s second-strike nuclear capabilities, thereby increasing preferences for first strike capabilities. Amoroso and Tamburrini [24] point out that even using AWS in a non-lethal manner to destroy buildings or infrastructure could have a global destabilising effect. Sharkey [16] also highlights concerns about global security due to an increase in the pace of war as a result of deploying AWS. In addition, he emphasises the likelihood of unpredictable interactions between different computational algorithms.

The fourth class of arguments against AWS, (iv above)—strongly connected to the first, and the inability to conform to IHL, are related to problems with the widespread use of algorithms in civil society and the demonstrated biases [25]. It is sometimes suggested that AWS could be used to pick out specific people, or classes of people, as legitimate targets of attack. However, it has become increasingly clear that decisions made about people using algorithms are frequently biased [26]. This is often the

result of problems with big data being used to ‘train’ machine learning systems. One problem is there has been a consistent failure to find a way to eliminate bias in the data. Moreover, machine learning algorithms are adaptive filters that smooth out the effects of outliers in the data, but these outliers could be minority groups that the trained system will consequently be less able to recognise. The resulting decision algorithms also lack transparency, since learning results in large matrices of numbers that are then used to generate the decisions. In civil society, this has proved to create bias in many domains such as juridical decisions, policing, mortgage loans, passport application and short-listing for jobs.

It is difficult to see how such algorithms could possibly be considered for a role in making decisions about who to target with automated weapons systems. This poses a particular problem for their use in counter-terrorism activities and border control. For example, face recognition algorithms are good at recognising the white males that form the majority of their training data and much worse at recognising black and female faces [27]. It is unlikely that the data used to train weapons systems to recognise particular individuals, or classes of people, will be trained with sufficiently representative sets of data to eliminate the problems of racial, gender and cultural biases.

## 4 Deception in Armed Conflict

In the US Army’s FM-90 manual, [28] there is clear recognition of the need for deception to achieve operational advantage. It provides an account of 10 maxims to be followed, which includes, ‘Cry-Wolf’ and ‘Magruder’s principles’. The ‘Cry-Wolf’ maxim represents the idea of desensitising the enemy to the likelihood of attack by repeated false alarms. For instance, in the week before the Pearl Harbour attack there had been 7 reports of Japanese submarines in the area, all of which turned out to be false. ‘Magruder’s principles’ refer to the exploitation of existing perceptions, and the notion that it is easier to strengthen an existing belief than to create a new one. In the D-day invasion, Hitler and his advisers were known to expect invasion in the Pas-de-Calais region, and efforts were made to strengthen this expectation.

The Joint-Publication JP3-13.4 [29] of the US forces identifies four basic techniques of military deception: (1) Feints (offensive action conducted to deceive the adversary about the location and or time of the main offensive action); (2) Demonstrations (a show of force without adversarial contact); (3) Ruses (deliberately exposing false or confusing information for collection by the adversary); and (4) Displays (simulation or disguising of capabilities which may not exist). As also discussed in that publication, military deception can involve electronic warfare. Electronic warfare has three major subdivisions: electronic attack (EA), electronic protection (EP) and electronic warfare support (ES). Camouflage and concealment are distinguished from military deception in JP3-13.4, although described as being able to support it by providing protection for activities.

Guides to deception for the military also refer to forms of deception that are against the laws of war, termed ‘acts of perfidy’, (Article 37 (1) Additional Protocol 1, [30]). Ruses of war; ‘acts which are intended to mislead an adversary or to induce him to act recklessly’, are not prohibited (Article 37 (2) Additional Protocol 1, [30]). Prohibited acts of perfidy are deceptions that lead the enemy to the false belief that they are entitled to, or are obliged to accord, protected status under the law of armed conflict. They include the use of vehicles marked with a red cross or a red crescent to carry armed combatants, weapons or ammunition, and the use in combat of false flags, insignia or uniforms. They are against the laws of war because they undermine the effectiveness of protective signals and jeopardise the safety of civilians and non-combatants.

Although some military guides distinguish between forms of deception, and camouflage and concealment, other accounts categorise camouflage as a form of deception [31]. Discussions of deception often hinge on the intentional deception of a human mind, or minds. In terms of the military, this might be the mind of the commander in charge of operations, or it might be the minds and perceptions of combatants on the ground. A slightly different account of deception arises when considering how AWS might be deceived.

The absence of human control of AWS necessitates a changed perspective on the notion of deception that has not yet made its way into military manuals. If, for instance, the sensors and programs of the autonomous weapons were subjected to deceptive strategies and, as a result, were to attack ‘friendly’ targets, or to plunge into the sea, this would not represent a deception of the human mind in a direct sense. As autonomous weapons, after they had been activated, they would have selected and attacked the targets without any human intervention. Should examples of AWS being disrupted in this way be described as deception? To answer this question, we need to re-examine what is meant by ‘deception’.

As we have already seen, the emphasis in military manuals and guides to deception of adversaries is on misleading the *mind or minds* of the enemy. But in the AWS examples above, human minds are not directly deceived. They could be said to be *indirectly* deceived, in that the operational commander’s intended target may not have been hit. But this is not the more straightforward version of deception assumed in the manuals. Is it still appropriate to use the term ‘deception’ here?

#### ***4.1 So, What Is Deception and Could a Weapon Be Deceived?***

Some definitions of deception require a person to have been deceived, and also that intention is involved. For instance, Carson [32] defines deception as ‘intentionally causing someone to have false beliefs.’ For him, deception can be distinguished from lying because deception implies success and that someone has been successfully caused to have false beliefs. A person who lies may not deceive the person to whom they lie. Zuckerman et al. [33], in a psychological investigation of deceptive communication, also define deception as requiring a human to have been deceived: deception

is ‘an act that is intended to foster in another person a belief or understanding which the deceiver considers to be false’.

Intentional deception can be undertaken with the aim of benefitting the deceiver. As well as examples of military deceptive strategies, there are others such as internet scams, or phishing attempts to gain information about someone’s bank details. Of course, it is also possible that a person or persons might intentionally deceive others with the aim of helping or improving their quality of life. Bok [34] gives several examples of deceptions created with good intentions, including placebos, and white lies.

Deception can also occur without intention as we have argued elsewhere [35]. Bok [34] points out various situations in which people might deceive without having intended to do so. They might deceive others by conveying false information in the belief that it is true. Deception also arises without intention in the natural world. In such cases, it is usually to the benefit of the deceiver. Bond and Robinson [36] define deception in the natural world as ‘a false communication that tends to benefit the communicator’. Examples include camouflage, mimicry, death feigning and distraction displays. Camouflage can make creatures less visible to their predators. Mimicry is used, for instance, by the edible viceroy butterfly which has the markings of the inedible monarch butterfly, and by the brood mimicry of the cuckoo. Death feigning as an anti-predator adaptation occurs in a range of animals, and distraction displays to draw attention away from nests and young are found in birds and fish. As Gerwehr and Glenn [37] point out, deception is used in the natural world ‘both to acquire dinner and to avoid becoming dinner’.

AWS could be used intentionally by humans in deceptive strategies. And, of course, programmed (or trained) computer algorithms do not have minds and thus cannot by themselves form an intention to deceive. But although they cannot intentionally deceive, they might be disrupted by deceptive strategies. We choose to describe AWS here as being deceived, despite our uneasiness about possible anthropomorphic language. In the present context, it is useful to use the term ‘deception’ as a shorthand to describe the situations in which the operations of AWS are disrupted, by either intentional or unintentional deception, and by deceptive strategies. Moreover, by saying that AWS can be deceived, we by no means wish to imply that they can be held responsible or accountable for their operations. At all times, responsibility for the behaviour of weapons rests with the humans who develop and use them (see e.g. [38]).

## 5 Deception and AWS

There are various ways in which AWS could be used to create a deception in the sense of some of the examples of military strategy described earlier. AWS are autonomous once launched, but the military can still be involved in decisions about when and where to deploy them. This is the case even if the weapons are set up to automatically launch when incoming missiles are detected—a decision has still been made by



humans to set them up in this way. AWS could be launched in an area as a *feint*: mounting an attack in an area to distract an adversary from an attack being prepared elsewhere. They could be launched as a *demonstration*, or show of force, attacking buildings or locations in order to create the impression of technological superiority. Of course, terrorists, non-state actors, and insurgents could also make use of AWS in a similar manner.

The more serious humanitarian risk is that deceptive strategies could be used against AWS to disrupt the behaviour of the machines. Humans are endlessly inventive and creative, and there is little reason to expect that the human targets of AWS will passively wait to be killed. Terrorists, insurgents, non-uniformed combatants and non-state actors are going to invent ways of deceiving and derailing AWS. Johnson [39] details the many adaptations and innovations of the Taliban in Afghanistan in the asymmetric warfare conducted there. Al Qaeda are known to make use of denial and deception strategies [40], and in 2013 Al Qaeda counter-drone manuals were discovered in Mali, detailing 22 steps for avoiding drone attacks [41]. Bolton [42] states ‘People are too messy, unpredictable, clever and tricky to meet the assumptions programmed into military technology’. He gives as examples the ways that Vietnamese communist soldiers spoofed the electronic detectors dropped from US warplanes onto their paths through the jungle: ‘they sent animals down the trail, placed bags of urine next to so-called “people sniffers”, and played tapes of vehicle noises next to microphones—prompting computerized bombers to unload explosives onto phantom guerrillas’.

It is easy to underestimate the technological ingenuity of low-tech actors. A good example was the US military capture of Shia militants with laptops containing many hours of video footage taken from US drones. They had used software called Skygrabber, available on the internet for \$26 dollars, for downloading music and video [43].

Hezbollah carried out similar operations against Israeli forces as far back as 1996 when they used photographic evidence taken from an Israeli drone of an Israeli attack. Hezbollah also claimed that they had used analyses of Israeli drone footage to plan ambushes, such as the “Shayetet catastrophe” in which 12 Israeli commandos were killed (the method they used to hack the drone signals remains unknown) [44].

An important motivation, for those likely to be subject to an attack from AWS, would be to cause them to select targets that reduced the risk of harm to either combatants or civilians. For instance, if they were deceived into attacking dummy buildings instead of military installations, expensive fire power could be drawn and exhausted. Similarly, it would be advantageous to find ways of camouflaging military targets from sensors such that they were shielded from attack. A deception that caused AWS to target neutral, or protected installations such as hospitals could create an effective public relations coup for a terrorist group (although it is not clear who should or would be held responsible in such a case). Of course, some forms of deception could have unwanted humanitarian consequences. For instance, if it was known that AWS were programmed to attack vehicles with the heat signature of tanks, efforts could be made to modify the tanks’ heat signature to resemble that of buses or lorries. But the unwanted consequence of this could be a subsequent modification to the AWS

sensors so that they targeted vehicles with the signature of buses or lorries, leading to wider devastation. This would be an example of the ‘monkey’s paw’ effect discussed in military accounts of deception, whereby seemingly effective deceptions result in unintended harmful side effects.

How could AWS be subject to deception? Once launched, they are dependent on their sensors and image recognition systems to detect the targets they have been programmed to attack. These are unlike human sensing systems and can be disrupted in ways that humans cannot even sense such as by means of high frequency sounds, bright lights, 2D images or even small dots that are entirely meaningless to us (see e.g. [45]).

There is growing awareness of the limitations of image recognition systems [46] and the risks that they could be unintentionally deceived or mistaken. For instance, problems with the sensors and image recognitions systems of autonomous cars have resulted in several Tesla crashes. Known objects in unexpected positions, such as a motorcycle lying on the ground, may not be recognised [47]. Self-driving cars and their sensors and software are known to have difficulties with rainy and snowy conditions [48]. In 1983, the sensors on Soviet satellites detected sunlight glinting on clouds and the connected computer system misclassified the sensor input as the engines of intercontinental ballistic missiles. It warned Lieutenant Colonel Petrov of an incoming nuclear attack—an unintentional deception [12].

Existing limitations are likely to be magnified by intentional efforts to mislead and confuse those sensors and image detection programs. The seemingly sophisticated sensors of AWS might be able to penetrate camouflage designed to fool human sensors. But, at the same time, available knowledge about the properties and limitations of the sensors used in computer control and classification could make it easy to hide from and misdirect AWS in ways that a human would not even notice.

There is a great deal of interest in the development of adversarial images. These are images that confuse image recognition systems trained using machine learning. For example, an adversarial image that to a human eye looks like a turtle can be perturbed by adding visual noise so that it is recognised by an image classification system as a rifle instead [49]. Adversarial images have been discussed in the context of autonomous cars—in one example, stickers added to a ‘Stop’ sign led to it being classified as a 45 mile speed limit sign [50]. There is research into ways of making classification systems resilient to adversarial images, for instance by training them to recognise them as adversarial, but it is not clear how successful this would be, and constant retraining would be needed to respond to new adversarial developments.

Another form of deception that risks AWS being directed to the wrong targets is ‘spoofing’ by sending a powerful false GPS signal from a ground base. This could cause AWS to mislocate and be guided to crash into buildings. Image classification systems could also be misled through the use of perfidious markers—such as placing a red cross on a military vehicle to prevent it from being targeted. In the light of adversarial images, it might be possible to mark such vehicles in a way that was picked up by an AWS yet was undetectable by the human eye. Other forms of perfidy are possible: for instance, if it were known that AWS would not target funeral processions, military manoeuvres could be disguised as funeral processions. Or if

they were programmed to avoid targeting children, combatants could walk on their knees.

Of course, human eyes might also be deceived by such disguises, but AWS are dependent on their programming, the limitations of the sensors and on the programmers having anticipated a deception from the infinite number possible in conflict. Humans can be susceptible to deception, but they also have an understanding of human social situations and would be able to interpret a social gathering in the way that a weapon could not. Also, in the fog of war, there is the possibility that humans might intuit that something was amiss. When Lieutenant Colonel Stanislov Petrov was on duty in 1983 at the Russian nuclear detector centre, he decided not to trust the repeated warnings from the computer about an incoming nuclear attack from the US. He did not initiate a retaliatory nuclear strike and correctly reported it as a false alarm [11, 12].

This problem of the lack of human intuition and understanding by AWS, robots, and computer systems stems from the same technological (or metaphysical) shortcomings as those that result in the inability to adhere to the principles of discrimination and proportionality. It is another manifestation of the limitations of programmed or trained algorithms. There is a risk that AWS will be misled by the inputs they receive, and that they will have already attacked before any human has had the opportunity to sense that something is not right and that a mistake is about to be made. The speed at which such weapons are likely to operate (an argument sometimes used in their favour) exacerbates this risk and means that even when humans see that something is going wrong, they will be powerless to stop them.

Another problem with AWS, mentioned earlier, is the danger of unpredictable interactions between different algorithms. As explained by Sharkey [11, 35], the algorithms controlling AWS will be kept secret from the enemy. That means that it is impossible to know what will happen when two or more top-secret algorithms from opposing forces meet each other. Apart from the unknown interactions, the algorithms could be programmed for deceptive strategies such as feinting or sensor disruption.

Not only are there reasons to fear the unpredictable effects of different algorithms interacting, there is also the problem of unexpected interactions between the programming of the AWS and unanticipated environmental situations. In software testing, it is well known that bugs and errors will remain in code. As programmed entities it is impossible to test AWS for their behaviour in all of the unanticipated circumstances that can arise in conflict. And it is impossible to ensure that their behaviour will not be catastrophic in an environment of deceptive strategies.

## 6 Conclusions

It is clear that a major weakness of autonomous weapons systems is that their sensors and image processing systems are vulnerable to exploitation for the purposes of deception. We argue that their application in the field would be subject to large scale

deception by enemy forces. Not only could the sensors that control the movement and target selection of AWS be misled through their limitations, their incoming information could be deliberately distorted by the enemy to alter attack strategies. Deceptions of AWS could result in wasted firepower, missed targets, and ‘friendly’ casualties and mishaps. The high speed at which AWS will operate and their autonomous nature, would make it difficult, perhaps impossible, for military commanders to prevent mistaken targeting even if they were to become aware of it. There is already a well-established set of arguments against AWS. Now add the risks of deception, and the impact that this would have on civilian populations and infrastructure, and the urgency is clear for an international legally binding prohibition treaty that comprehensively bans the development, production and use of weapons that operate without meaningful human control.

## References

1. Tzu S (2018) *The art of war*. Translated by Lionel Giles, Benediction Classics (original 5th Century BC)
2. Abbot J (1901) *Hannibal*. Harper and Brothers Publishers, New York and London. Accessed <https://www.heritage-history.com/index.php?c=read&author=abbot&book=hannibal&story>
3. Tucker SC (2013) Editor: *American civil war: the definitive encyclopedia and document collection*, p 1587
4. Pope D (2005) *The battle of the river plate: the hunt for the German pocket battleship Graf Spee*. McBooks Press
5. Garber M (2013) *Ghost army: the inflatable tanks that fooled Hitler*. The Atlantic, 22 May
6. Brown A (2007) *Bodyguard of lies: the extraordinary true story behind D-day*. The Lyons Press
7. US Department of Defense (2012) Directive 3000.09. In: *Autonomy in Weapons Systems*, 21 November, pp 13–14
8. Human Rights Watch (2012) *Losing humanity: the case against killer robots*. Accessed 17 May 2018. Available at: <http://www.hrw.org/reports/2012/11/19/losing-humanity-0>; Internet
9. ICRC (2014) *ICRC, autonomous weapon systems: technical, military, legal and humanitarian aspects*. In: *Expert meeting, vol 1, Geneva, Switzerland, 26–28 Mar 2014*
10. Amoroso D (2020) *Autonomous weapons systems and international law: a study on human-machine interactions in ethically and legally sensitive domains*. Edizioni Scientifiche Italiane, Napoli
11. Sharkey N (2020) *Fully autonomous weapons post a unique dangers to human kind*. Scientific American, February
12. Scharre P (2018) *Army of none: autonomous weapons and the future of war*. W.W. Norton and Company
13. Crootof R (2015) *The killer robots are here: legal and policy implications*. *Cardoza L Rev* 36:1837
14. Zacharius G (2015) *US armed services committee hearing on “advancing the science and acceptance of autonomy for future defence systems”*. <http://armedservices.house.gov/index.cfm/2015/11/advancing-the-science-and-acceptance-of-autonomy-for-future-defense-systems>
15. *US Mission Statement (2020)* <https://geneva.usmission.gov/2020/09/30/group-of-governmental-experts-on-lethal-autonomous-weapons-systems-laws-agenda-item-5d/>
16. Sharkey N (2012) *The evitability of autonomous robot warfare*. *Int Rev Red Cross* 94(886):787–799
17. Suchman L (2016) *Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy*. In: *Panel presentation at CCW informal meeting of experts on lethal autonomous weapons, Geneva, 12 April 2016*

18. Sharkey A (2017) Can we program or train robots to be good? *Ethics Inf Technol* (Online First). <https://doi.org/10.1007/s10676-017-9425-5>
19. Heyns C (2017) Autonomous weapons in armed conflict and the right to a dignified life: an African perspective. *South Afr J Hum Rights* 33(1):46–71
20. Asaro P (2012) On banning autonomous lethal systems: human rights, automation and the dehumanizing of lethal decision-making, special issue on new technologies and warfare. *Int Rev Red Cross* 94(886, Summer 2012):687–709
21. Sharkey A (2019) Autonomous weapons systems, killer robots and human dignity. *Ethics Inf Technol* 21(2):75–87
22. Tamburrini G (2016) On banning autonomous weapons systems. From deontological to wide consequentialist reasons. In: Bhuta N et al (eds) *Autonomous weapons systems. Law, ethics, policy*. Cambridge University Press, pp 121–141
23. Sharkey N (2008) Grounds for discrimination: autonomous robot. *RUSI Defence Syst* 11:86
24. Amoroso D, Tamburrini G (2017) The ethical and legal case against autonomy in weapons systems. *Global Jurist* 17:3
25. Sharkey N (2018) The impact of gender and race bias in A.I. *Humanitarian Law Policy Blog* August 2018
26. O’Neil C (2016) *Weapons of math destruction: how big data increases inequality and threaten democracy*. Penguin Books
27. Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In: *Proceedings of machine learning research, 2018 conference on fairness, accountability, and transparency*, vol 81, pp 1–15
28. Field Manual FM 90-2 (1988) *Battlefield deception*. US Army Washington DC
29. Joint Publication 3-13.4 (2006) *Military deception*. Joint Chiefs of Staff, USA
30. Article 37 (1977) Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (Protocol I), 8 June 1977
31. Field Manual FM 3-13.4 (2019) *Army support to military deception*. <https://armypubs.army.mil>
32. Carson TL (2010) *Lying and deception: theory and practice*. Oxford University Press Inc., New York
33. Zuckerman M, DePaulo BM, Rosenthal R (1981) Verbal and nonverbal communication of deception. In: Berkowitz L (ed) *Advances in experimental social psychology*, vol 14. Academic Press, New York, pp 1–59
34. Bok S (1999) *Lying: moral choice in public and private life*. Second Vintage Books Edition, New York
35. Sharkey A, Sharkey N (2020) We need to talk about deception in social robotics! *Ethics Inf Technol* (Published online 11 November)
36. Bond CF, Robinson M (1988) The evolution of deception. *J Nonverbal Behav* 12(4):295–307
37. Gerwehr S, Glenn RW (2020) *The art of darkness: deception and urban operations*. RAND Corporation, MR-1132-A, 2000, Santa Monica, Calif. As of July 20, 2020. [https://www.rand.org/pubs/monograph\\_reports/MR1132.html](https://www.rand.org/pubs/monograph_reports/MR1132.html)
38. Johnson DG, Powers TM (2005) Computer systems and responsibility: a normative look at technological complexity. *Ethics Inf Technol* 7(2):99–107. <https://doi.org/10.1007/s10676-005-4585-0>
39. Johnson TH (2013) Taliban adaptations and innovations. *Small Wars Insurgencies* 24(1):3–27
40. Jessee DD (2006) Tactical means, strategic ends: Al Qaeda’s use of denial and deception. *Terrorism Polit Violence* 18:367–388
41. Burgers TJ, Romaniuk SN (2017) Learning and adapting: Al Qaeda’s attempts to counter drone strikes. *Terrorism Monit* 15:11
42. Bolton M (2020) New book shows catastrophic folly of automating warfare. In: *Blog on ICRAC (international committee for robot arms control) website*, posted September 20th 2020
43. MacAskill E (2009) <https://www.theguardian.com/world/2009/dec/17/skygrabber-american-drones-hacked>

44. Defensetech (2010) <https://www.military.com/defensetech/2010/08/10/hezbollah-claims-it-hacked-israeli-drone-video-feeds>
45. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A (2016) The limitations of deep learning in adversarial settings. In: 2016 IEEE European symposium on security and privacy (EuroS P), pp 372–387. <https://doi.org/10.1109/EuroSP.2016.36>
46. Cummings ML (2020) Rethinking the maturity of AI in safety critical settings. In: Advances in artificial intelligence magazine
47. Alcorn MA, Li Q, Gong Z, Wang C, Mair L, Ku WS, Nguyen A (2018) Strike (with) a pose: neural networks are easily fooled by strange poses of familiar objects. arXiv: 1811.1153
48. Zang S, Ding M, Smith D, Tyler P, Rakotoarivelo T, Kaafar MA (2019) The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog and hail affect the performance of a self-driving car. IEEE Veh Technol Mag 103–111
49. Athalye A, Engstrom L, Ilyas A, Kwok K (2018) Synthesizing robust adversarial examples. In: Proceedings of the 35th international conference on machine learning (PMLR 80: 2840293)
50. Eykhold K, Evtimov I, Fernandes E, Li B, Rahmati A, Xi C, Prakash A, Kohno T, Song D (2019) Robust physical world attacks on deep learning models. arXiv: 1707.08945

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Weapons of Mass Destruction—Conceptual and Ethical Issues with Regard to terrorism



Jonas Feltes

## 1 Introduction

The concept of WMD is part of numerous national laws and is the core of one of the most important treaties of the United Nations [51, 64]. Yet, the definition of what should be considered a WMD is far from established and subject to controversial debates. Academics, policymakers, and legislators have been introducing a variety of partly conflicting conceptualizations of WMD into scientific debates, public discourse, and legislations over the last eight decades. Hence, it is unsurprising that this concept and its changing definition have been subject to politicization. Especially in light of the so-called “War Against Terror,” WMD became the synonym of a worst-case terrorist attack scenario that ought to be prevented by any means [55]. However, terrorism and other asymmetrical conflicts pose serious challenges to the concept of WMD—serious enough to think about alternatives to this term in case of counter-terrorism discussions. One particular issue stems from the ethical challenges that the label WMD generates if used in combination with terrorism.

This chapter presents the history of the term WMD as well as numerous issues with and alternative approaches to the concept of WMD. In this discussion the concept of CBRN (chemical, biological, radiological, and nuclear) as a prominent interpretation of WMD is of utmost importance. It will be argued that a static concept that includes or excludes certain weapon types purely on the basis of their physical impact in an attack deals with problematic threshold issues and ethical challenges. In this chapter, I discuss concepts of terrorist weaponry that are focused on a more complex account of the impact of each weapon type used by terrorists. Specifically, the impact of

---

This chapter is part of the author’s doctoral dissertation titled “CBRN Threats, Counter-Terrorism, and Collective Moral Responsibility”.

---

J. Feltes (✉)  
TPM Faculty, TU Delft, Delft, The Netherlands  
e-mail: [j.feltes@tudelft.nl](mailto:j.feltes@tudelft.nl)

© The Author(s) 2021  
A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,  
Advanced Sciences and Technologies for Security Applications,  
[https://doi.org/10.1007/978-3-030-90221-6\\_4](https://doi.org/10.1007/978-3-030-90221-6_4)

a weapon type will be assessed by means of analyzing its hard (physical) and soft (psychological, economic, political) damage. Furthermore, the time that is necessary to create a high impact with the one-off use of the weapon, as well as uncertainties with regard to the consequences of the use of said weapon, will be part of the impact assessment.

However, in order to assess the dangers involved in and the severity of specific weapons in the hand of terrorists, it is not sufficient to focus only on the impact of a possible attack with this weapon. For example, even without an elaborate analysis, it is clear that nuclear weapons would easily achieve the highest score in terms of impact. However, the impact of a certain weapon technology does not say much about the terrorist threat posed by this weapon if this technology is simply not available to terrorist groups. Hence, a basic assessment of the resources and other restricting factors that guide the weapon choices of terrorists needs to be part of this chapter as well. This assessment might show a trend that is diametrically opposed to the impact of specific weapon technologies. It includes, for example, factors like accessibility, required expertise, operational space needed as well as tactical advantage and ideological considerations.

With this more complex understanding of the impact of a certain weapon and its availability to terrorists, the threat that terrorist attacks with improvised unconventional weapons can be analysed and displayed more accurately. This does not only allow for more efficacious and precise countermeasures, but also reduces ethically unsustainable behaviour of first responders and the press during a terrorist incident.

## 2 The (Never-Ending) History of WMD and CBRN

The notion of weapons of mass destruction has its origins in the middle of the twentieth century. One of the first recorded uses of the term WMD dates back to 1937 when the Archbishop of Canterbury warned against “all the new weapons of mass destruction” during his Christmas address [13], pp. 6–8. The archbishop never specified what kind of weapons he referred to in his address. Yet, researchers have been arguing that the term and the address, in general, was designed as a response to the bombing campaigns against civilians in Spain and Asia during that year [13], pp. 6–8. However, as Seth Carus argues, the Archbishop was also actively concerned with novel weapon systems like chemical warfare and could very well have referred to chemical or even biological weapons with the term weapons of mass destruction [13], p. 7.

The first politically relevant and precise notion of WMD was delivered roughly eight years after the Christmas address of the Archbishop of Canterbury. On 15 November 1945, the political leaders of the United States, Canada, and the United Kingdom issued a joint declaration calling for the regulation of atomic energy. In this declaration, the authors called amongst others “[f]or the elimination from national armaments of atomic weapons and of all other major weapons adaptable to mass destruction” (opp. cit. Carus [13], p. 8). An even more precise notion of WMD was



defined only three years later by the United Nations Commission on Conventional Arms Control (CCA). The CCA issued an official definition of WMD and characterized this concept as chemical, biological, radiological, and nuclear (CBRN) weapons. Furthermore, the CCA opened up this definition towards potential, novel weapon systems “which have characteristics comparable in destructive effect to those of the atomic bomb or other weapons mentioned above” (opp. cit. Carus [13], pp. 9–10).

Another important part of the history of WMD and CBRN is the strategic use of the term WMD for political ends. As Michelle Bentley shows in a convincing argument, WMD has been defined and interpreted in different ways by different political actors in order to further political agendas (See Bentley [8, 9]). For example, the U.S. government and specifically the Department of Defense (DOD) appeared to favor a definition of WMD that exclusively refers to CBRN devices that are *capable of mass destruction*. Note that this definition would potentially exclude low-yield nuclear devices. As Bentley argues and Carus suggests, this slightly different—and ambiguous—definition had political advantages for the USA [8], pp. 392–393, [13], p. 31. Amongst others, it would enable the U.S. military to deploy low-yield nuclear weapons in space or the deep sea, although the UN Space Treaty and the Sea Bed Treaty prohibited the deployment of WMDs in space or the deep sea. Because of these changing definitions of WMD that admittedly only differed in nuances from the CBRN-based understanding of WMD, Bentley argues that WMD should be understood as a non-essentialist term rather than as a static definition. Furthermore, Carus managed to identify six different understandings of WMD in national and international discourses, of which most are based on (some) CBRN technologies [13], p. 36. The most controversial interpretations of WMD in this list (such as weapons of mass effect) will be discussed below.

### 3 Abandoning WMD Altogether?

Researchers have identified several different problems with the concept of WMD that range from conceptual issues to implementation issues in intelligence and law enforcement practice. In particular, Christian Enemark has been stressing the problems of the term “WMD”. In a pivotal article for this discussion, Enemark states:

“The WMD label exaggerates the destructiveness of chemical weapons, misrepresents the problem of biological weapons, and diverts attention from the overriding importance of dealing with nuclear weapons” [25], p. 382.

This heterogeneity of weapon types summarized under the umbrella term of WMD certainly poses challenges to the concept of WMD. These challenges are even more pressing when dealing with improvised CBRN weaponry. As past incidents of use of chemical agents showed, attacks using chemical or even radiological weapons do not inflict mass casualties comparable to those casualty numbers expected for the deployment of, for example, a nuclear weapon or a weaponized biological agent (For

cases see Danzig et al. [18]; The Times of Israel [63]). In fact, researchers have argued that, for example, improvised radiological weapons do not produce more physical impact than IEDs or other conventional weapons [36], p. 73.

Moreover, even each of the four major weapon types summarized under the term WMD seems too broad to account for terrorist weapon technologies. For example, the use of salmonella bacteria to terrorize innocent people would certainly count as improvised biological warfare but does not create the devastating consequences that a weaponized Marburg virus may be capable of. The salmonella campaign of the Rajneesh cult in 1984 is a case in point here [24], p. 59. Thus, it seems inaccurate to refer to all CBRN weapons as “weapons of mass destruction”. The extent of destructiveness between these four categories, but also within each of these categories, is too diverse to group all of these weapons under the term WMD.

However, contrary to Enemark’s position, one could think of at least three different arguments against the radical abandonment of WMD: First of all, it is simply impossible (and undesirable) to remove the concept of WMD from international law and diplomacy. Seth Carus shows in a detailed analysis that the term of weapons of mass destruction is an essential concept in many of the most relevant international treaties including the Chemical Weapons Convention (CWC), Biological Weapons Convention (BWC), the Nuclear Non-proliferation Treaty, the Strategic Arms Reduction Treaty (START), the Space Treaty, and the Seabed Treaty [13], pp. 6–34. Abandoning the term WMD would mean to, potentially, having to jeopardize or even renegotiate these treaties.<sup>1</sup> Secondly, Bentley points out in a well-crafted argument that the term WMD is a non-essentialist concept that is being re-defined and used by political actors in order to further political agendas. This active role of WMD as a strategic tool in politics makes it almost impossible to abandon it from policymaking (See Bentley [8]). Lastly, it should be noted that military-grade biological, chemical, and nuclear weapons that are stockpiled and deployed by nation-states have common characteristics that could make the WMD concept useful for military strategists: For example, all three weapon categories require decontamination and extensive protective gear and all three weapon categories include strictly anti-personnel capabilities that outperform the blast radius of conventional weapons.

Yet despite the arguments in favor of keeping WMD as a concept in general, one still has to account for Enemark’s criticism of diversity of impact within this concept. One possible solution would be to adopt the strongest definition of WMD as presented in Carus’s article that only classifies those CBRN weapons as WMD that are, in fact, mass destructive [13], p. 36. Obviously, this classification almost immediately poses a threshold level problem: what should be considered mass destruction in this regard? One way of arguing would be to favor a *potential* mass destructiveness of certain CBRN weapons: while a nuclear warhead, the Novichok virus or a weaponized Marburg virus could potentially kill thousands of people in a one-off use, Salmonella

---

<sup>1</sup> Enemark argues against this by stating that WMD is a redundant term in international treaties that could be simply replaced by chemical, biological, or nuclear weapons. However, as Bentley has shown, the term WMD is more than a summarizing term of NBC, but a political tool. Because of this historically grown relevance of the term, it might, in fact, not be as easy to replace it in international treaties as Enemark suggests (See [8, 25, 26]).

bacteria or a dirty bomb are not capable of doing so. Obviously, this interpretation of WMD is not flawless as it allows certain strategic and politically motivated exclusions or inclusions to the WMD category, as seen above. However, in light of Enemark's strong case against the concept on the one hand and good reasons to keep WMD on the other, the definition of WMD as military-grade CBRN weapons that have been in national military arsenals at some point and that are actually capable of mass destruction seems to be the least problematic choice and will be used in the next section of this chapter.

## 4 WMD and Terrorism

It is important to note that, despite massive amounts of WMD-related research and threat assessments in terrorism studies,<sup>2</sup> WMDs (defined as military-grade CBRN weapons with mass destructive effects) are almost absent in the arsenal of the most relevant terrorist groups. Yet, not only WMDs, but even the use of the much broader weapon group of CBRN weapons in general (mass destructive or not) seems to be the exceptional more than the rule in terrorism. According to the Global Terrorism Database (GTD), the most comprehensive collection of terrorist incidents, only 0.233% of all recorded terrorist attacks were committed with CBRN weapon technologies. The majority of these cases were targeted poisonings and the use of CS or tear gas [60]. Based on an empirical assessment of terrorist attacks against the United States of America, the authors of another study note that “[b]etween 1970 and 2010, there were 751 terror attacks using conventional explosives and only 85 attacks using CBRN weapons” [24], p. 58. Moreover, the authors of this study have included very low-impact CBRN incidents such as attempted poisonings.

Furthermore, the concept of WMD, as defined above, does not encompass all mass destructive terrorist events or all terrorist weapons of mass destruction. Indeed, many of the past terrorist attacks that produced exceptionally large amounts of fatalities were executed with weapons that would not qualify as WMD as defined above. The attack on September 11, 2001, in New York City is just one (prominent) example of such weapons (See discussion in Bentley [8], p. 397). Furthermore, it has been shown in different studies that the most deadly terrorist attacks have been committed with conventional weapons such as IEDs or firearms. For instance, the authors of the recent studies on WMD terrorism in the USA that was mentioned above note in this regard:

In addition to their higher attack frequency, conventional attacks using explosives cause higher damage, on average (...) Since 1970, 216 people have died from terrorist bombings in the USA while seven individuals have died from CBRN attacks. On average, 0.28 people die per bombing campaign, while 0.08 people die per CBRN attack [24], p. 59.

In addition to this observation, a quantitative data analysis of the incidents listed in the GTD calculated both the total numbers of fatalities as well as the fatalities per

---

<sup>2</sup> A brief selection of published research includes [2–5, 8–10, 13, 14, 25, 26, 36, 37, 39, 45, 53, 54].

attack for different weapon types used by terrorist groups (See LaFree et al. [46]). Based on this calculation, vehicle-based attacks seem to be the deadliest terrorist weapons, followed by melee weapons and firearms. According to this study, chemical weapons come in fourth and are the deadliest weapons that are commonly considered WMDs—with a total fatality number of 629. In comparison, explosive devices have a slightly lower rate of fatalities per attack but are responsible for a total amount of 99,379 deaths [46], p. 139.

Because of the absence of WMDs in terrorist incidents, one could argue that this weapon category should not have priority and should not be discussed to such an extent in terrorism research. However, next to the low probability that a terrorist group, in fact, gets their hands on a WMD, law enforcement and security agencies have been using the term WMD with regard to terrorism to stress the danger of certain non-CBRN weapons with particularly high impact. In these instances, the notion of mass destruction has arguably lowered threshold levels when referring to crimes or terrorism in comparison to the above-formulated definition of WMDs as military-grade CBRN weapons. Even a death toll in the lower hundreds caused by an improvised device could count as a WMD event in the eyes of practitioners and policymakers:

In the USA, this approach to redefine WMD for terrorism was even turned into national legislation. In the aftermath of the Oklahoma City bombing in 1998, the perpetrator of the attack, Timothy McVeigh, was sentenced to death in accordance with a by then only one-year-old reform of the US criminal code (For discussion, see Madeira [49]). According to these changes, the use of a WMD can be punished with the death sentence and WMD in this regard does not only refer to CBRN devices, but also to other “destructive devices include[ing] bombs, grenades, mines, or any gun with a barrel larger than one-half inch” (opp. cit. Carus [13], p. 29). In this reform, the term WMD does not only refer to CBRN weapons, but could better be characterized as CBRNE (chemical, biological, radiological, nuclear, and explosive). Next to Timothy McVeigh, also the shoe bomber Richard Reid as well as the perpetrators of the Boston Marathon bombing were prosecuted for using WMDs—despite the fact that all these attacks involved conventional IEDs.

The interpretation of WMDs as CBRNE is one of the most prominent proposals to cope with the challenges of the concept of WMD with regard to terrorism. Next to practical and legislative advantages, the interpretation of WMD as CBRNE in terrorist incidents also appears to be a solution to the problem that the above-defined interpretation of WMD as military-grade CBRN may be both too narrow and factually irrelevant to account for most mass-casualty terrorist attacks. By adding explosive weapons, that were used in 52.65% of all terrorist attacks listed in the GTD [60], the concept of WMD rapidly becomes a synonym for the most worrisome and most destructive weapons in terrorism—as the term traditionally promised.

Despite these obvious advantages, the treatment of WMD as CBRNE extrapolates some of the problems Enemark is raising in his article. For example, the problem that WMD includes too diverse weapon types that cannot be summarized in a single category becomes even more severe with regard to the CBRNE interpretation. The addition of explosive weapons to the definition of weapons of mass destruction

would further broaden the concept and would, for example, refer to the nuclear bomb and to small IEDs that contain little more than pyrotechnical substances alike. Furthermore, if one would interpret explosive weapons as not only referring to IEDs but also to RPGs, mortars, grenades, and small artillery, then the category of WMD would include almost all known weapon types with the exception of small firearms and melee weapons. This interpretation of WMD seems to be too broad to be an efficacious category for both symmetrical and asymmetrical conflicts. Efficacious in this regard does not only mean that the CBRNE interpretation of WMD seems too diverse from a theoretical perspective.

It also poses serious challenges for the practitioners and institutions that work with this definition. First of all, the CBRNE definition fundamentally conflicts with the definition of WMD used in international law and numerous UN regulations and treaties. Furthermore, since the label CBRNE presents itself as a single category of (advanced) weaponry, law enforcement, and intelligence practitioners could be tempted to allocate a special branch of their work to this category. However, since the weapons summarized under this label are highly diverse, some of them need completely different resources and analysis than others. For example, counter-measures against nuclear terrorism ought to focus on global non-proliferation efforts and state-funded terrorism, while IED counter-measures are (amongst others) focused on restricting access to certain household chemicals. The CBRNE label could be falsely suggesting that the threats evolving out of these different weapon types should be treated within the same department or group of analysts.

Moreover, and relevantly for this chapter, the interpretation of WMD as CBRNE with regard to terrorism poses some serious ethical issues that can be portrayed with the help of two examples.

On June 12, 2018, German security forces stormed an apartment in Cologne and arrested the Tunisian Salafist Sief Allah H. on the basis of intelligence that he planned a terrorist attack. During the raid of his apartment, Special Forces were called in and found over 3000 castor beans that contain the organic toxin ricin. According to the German police report of this incident, Sief Allah H. had already begun to grind the seeds and had apparently attempted to combine the ricin powder with an IED (improvised explosive device) to disperse the toxin in a populated area in Cologne [56, 59, 61].

In the aftermath of this plot, Sief H's plan to construct a ricin-based IED was portrayed as a singular and exceptional case of terrorism that had the potential to kill or wound tens of thousands of persons. H's device was repeatedly called the first "bio bomb" in the history of terrorism in Germany [19]. This characterization of the incident that the German news media called the "Cologne Ricin Plot" fits all too well into the above decided interpretation of all CBRNE weapons as WMDs.

This portrayal of the ricin plot as WMD plot was visible in the journalistic reporting on the incident. In many journalistic analyses of the plot, authors described ricin as a biological weapon agent and referred to the Chemical Weapons Convention (CWC) and in the Biological Weapons Convention (BWC) of the United Nations (UN) [51, 64, 67]. This interpretation of the plot as a WMD event significantly influenced the style of reporting in the German news media. After the arrest of Sief

Allah H., the German daily newspaper *Rheinische Post* published an article about the details of H's plot. In the title of this article, the author claimed that the amount of ricin that H. produced had the potential to kill up to 13,500 persons [56]. Although German counter-terrorism forces managed to arrest H. before he could commit the attack, the journalist reporting on the incident and the hypothetical scenarios that were formulated in the headline of the article, arguably, evoked a substantial amount of anxiety among the German public. Furthermore, one could argue that this style of reporting contributed to an erosion of public trust in the German security apparatus. The mere prospect of an attack with up to 13,500 fatalities was more than enough to spread fear and distrust in German society.

In the text of the article in *Rheinische Post*, the author explains that the estimate of 13,500 potential fatalities on the basis of the ricin in Sief H.'s apartment was given by a German security official. However, the author admits in a short sentence that the same official also stated that the number of 13,500 was a mathematical calculation on the basis of the LD<sup>50</sup> value<sup>3</sup> of ricin [56]. Yet, the LD<sup>50</sup> value exclusively displays the lethality of a perfectly purified substance under ideal laboratory conditions. Later on in the article, the author, in effect, admits that this was an exaggeration when he stated that the interviewed security official estimated the lethality of H.'s actual ricin device to be in the low hundreds. While this death toll would still be horrific, it would not be the almost apocalyptic number of 13,500 fatalities after a single attack, as was propagated in the title of the article.

The security official that was interviewed for the article gave a differentiated estimate of the possible consequences of an actual attack with H.'s device. Yet, apparently, this estimate was not in line with the picture of a planned WMD attack that the journalist wanted to communicate with the article. Hence, he chose to use the estimate that was based on the LD<sub>50</sub> value of ricin as the headline of the article. However, with this headline, the article clearly provided H. with the means to greatly increase fear among the German public. This fear was in large parts generated by the WMD label that was pinned to the Cologne Ricin Plot.

The ethical issues that arises here stem from the coverage of the plot as one of WMD: In presenting H's plot as one of WMD with the potential to cause 13,500 fatalities, the press coverage likely spread significant fear through the relevant population. The issue here is that, unknowingly, the press coverage may cause caused 'soft damage', where a particular attack has population level psychological impact by means of causing widespread fear in society, a point returned to below. The issue here is that presenting the Ricin plot as one of WMD in fact aided the social impacts that H might have been seeking. Thus we have an ethical issue about responsibility for soft damages, and how we ought to assign that responsibility to actors other than the terrorist themselves.

Another example of these ethical challenges is the 2006 Forest Gate Raid, which was based on intelligence provided by the British intelligence agency MI5 that a

---

<sup>3</sup> The LD<sub>50</sub> value refers to the lethal dose of a substance and describes how many µg (or mg) per kg body weight of the substance is necessary to kill 50 percent of the exposed population under laboratory conditions.

radiological or chemical device was stored for an attack in two apartments in a neighbourhood of London [11]. However, during the raid this piece of intelligence turned out to be false. In fact, the two residents of the raided apartment did not have any ties to terrorism. Yet, not knowing about this false intelligence, the police arrested the residents of the apartment and one of the officers shot a resident in the chest (Independent Police Complaints Commission 2006).

Here, the anticipated, devastating consequences of a ready-to-use WMD coupled with the full-body protective gear that influenced the officer's sensory apparatus and further heightened the stress associated with the threat caused the officer to shoot the resident. As a response to this, the British prime minister Tony Blair commented on the raid as follows: "You can only imagine if they [police officers] fail to take action and something terrible happened what outcry would be then, so they are in an impossible situation" [6]. This raid was an "impossible situation" for the operatives (and the residents) since the time pressure and the stress of an already assembled WMD with its anticipated consequences forced quick response and caused mistakes and overreaction.

This has ethical relevance, as counting all CBRNE weapons as WMDs in a terrorist attack can falsely extrapolate the gravity of the situation that police officers on site might be confronted with. In case of the Forest Gate Raid, the police officers entered the apartment with the expectation to be forced to prevent an attack of tremendous destructive potential at all costs. The (implicit) labelling of all radiological or chemical devices as WMDs caused the police officers to act disproportionately.

Here the ethical implications are twofold: first, using a coarse and broad definition of WMD means that the security officials themselves perceive a particular operation as posing significant risk to them. This places unjustified stress and pressure on those security officials, which leads to the second ethical issue—in engaging with a potential target as not simply a terrorist, but one with potential WMDs, it is more than likely that the counter-terrorism response will be disproportionate to the actual objective threat that they are facing. Proportionality is a fundamental ethical principle for security actors, and so we need to be very careful with the use of terminology like WMD that might both induce and potentially be seen to justify a disproportionate response to the actual threat being faced.

Both of the above-described examples show that CBRNE definition of WMD poses serious ethical challenges in practice and is still focused on physical impact as a defining criterion. However, as will be shown below, the impact of a weapon in the hands of terrorists should not only be characterized by focusing on its capability to produce mass physical destruction. Several authors pointed out that the impact of a terrorist weapon consists of multiple different categories including, but not limited to, physical destructiveness (See e.g. Bunker [12], Dunn et al. [23]). Selected approaches to give alternative concepts to classify especially impactful terrorist weapons will be discussed in the following section.



## 5 Alternative Concepts for Terrorist Weapons of Mass Destruction

The issues associated with mass casualty terrorist events and the definition of WMD caused several researchers, practitioners, and policymakers to rethink the conceptualization of terrorist weaponry.

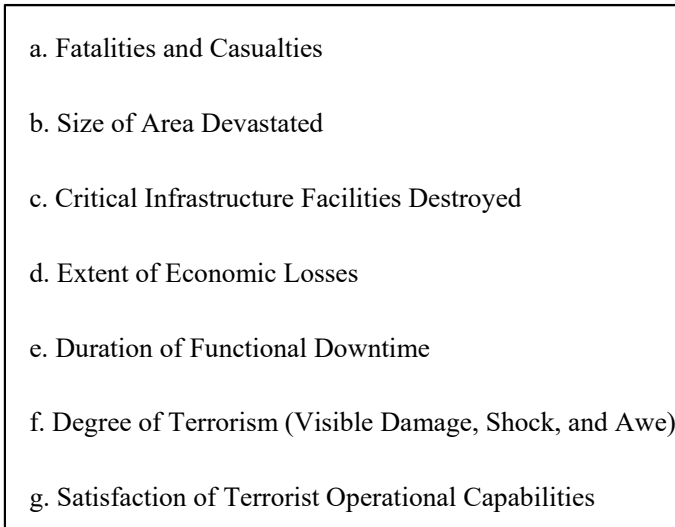
One possible solution to the problem of defining WMD was proposed by Robert J. Bunker, who presented his concept of Weapons of Mass Disruption (WMD<sup>2</sup>) in a publication in 2000 (See Bunker [12]). In his article, Bunker points out that certain novel weapon types (including CBRN weapons like non-lethal viruses) cannot be classified as causing mass destruction. Bunker argues that these weapons target relationships and bonds on a massive scale (mass effect) in society rather than physical objects and persons [12], pp. 41–43. Therefore, these weapons might have an enormously disruptive effect despite not inflicting mass casualties or large-scale physical destruction. Clearly, Bunker's novel concept of WMD<sup>2</sup> could be used to solve the problem that some WMDs such as radiological dispersal devices (RDDs) do not seem to be mass destructive, but rather mass *disruptive* in societies. However, in solving this problem, Bunker creates yet another category of weapons that is arguably as vague as WMD. The concept of WMD<sup>2</sup> does not seem to have clear borders and threshold values with regard to effect size and extent of disruption. Thus, Bunker's solution to the problems of WMD creates even more problems with regard to vagueness and fuzzy borders between weapon categories. Furthermore, many of Bunker's examples of WMD<sup>2</sup> weapons (i.e., radio frequency weapons, genetic alteration weapons, liquid metal embrittlement) seem even more detached from the reality of terrorist weapon choices than the traditional WMD weapon category.

Perhaps the most promising candidate concept in relation to mitigating the problems of WMD with regard to terrorism is the concept of Weapons of Mass Effect (WME). Initially proposed by William Yengst in 2008, the concept of WME is aimed at accounting for all those (terrorist) weapons that cannot be considered strictly mass destructive in the traditional sense, but that create a mass effect (See Yengst in Dunn et al. [23]). Yengst defines mass effect as an interplay of seven different criteria (Fig. 1):

According to Yengst, these criteria can be used as a rating system for terrorist weapons: only if a particular weapon reaches a certain score with each of these criteria and surpasses a certain threshold (in Yengst's analysis 41 points), then one could reasonably call this weapon a weapon of mass effect. Examples of these WMEs in Yengst's analysis are explosive attacks against critical infrastructure, the use of kinetic energy against office buildings (e.g., with an aircraft) or the contamination of drinking water supplies. With his approach to a dynamic rating system, Yengst effectively circumvented the demarcation problems resulting from static concepts such as WMD. Thereby, he solves problems such as the lacking identification of mass destruction and the high diversity of weapon types within the concept of WMD.

However, Yengst's proposal of WME does not abolish or replace the concept of WMD but rather offers an additional category of terrorist weapons for all those



- 
- a. Fatalities and Casualties
  - b. Size of Area Devastated
  - c. Critical Infrastructure Facilities Destroyed
  - d. Extent of Economic Losses
  - e. Duration of Functional Downtime
  - f. Degree of Terrorism (Visible Damage, Shock, and Awe)
  - g. Satisfaction of Terrorist Operational Capabilities

**Fig. 1** Yengst's criteria for mass effect [23], pp. [2–5] 4–5

unconventional weapon types that are not regarded WMDs in the traditional sense. While the dynamic nature of Yengst's approach does not run into the same problems as Bunker's WMD<sup>2</sup> proposal, it does not explicitly solve the problems with the concept of WMDs, since only a few of Yengst's WME examples challenge the concept of WMD. Furthermore, Yengst's concept of WME allows for a large degree of subjectivity concerning the presumed effect of a weapon or an attack. For example, a workshop report from 2010 that used Yengst's concept portrayed the 9/11 attacks, the Popular Front for the Liberation of Palestine aircraft hijackings in the 1970s as well as the attempted assassination of Margaret Thatcher with an IED as WME attacks.

To sum up, while Yengst's approach to introduce a rating system to measure the effect (or impact) of terrorist weapons appears to be a suitable candidate to resolve a number of the problems with the concept of WMD in relation to terrorism, his introduction of the static WME concept for high-scoring weapons re-introduces some of these problems. By including or excluding certain weapon types to this concept according to varying criteria, researchers that use WME are yet again facing the problems that have been discussed above with regard to WMD. Hence, and based on Yengst's proposal, the following section will propose to expand Yengst's idea of a rating system to measure the impact of terrorist weapons. However, contrary to Yengst's approach, this new proposal does not introduce yet another concept of high-impact weapons but rather treats each and every (potential) terrorist weapon individually and based on its score in the rating system.

## 6 The Terrorist Weapon Rating System

As seen in the last section, some researchers and practitioners have made attempts to overcome the problems arising from the traditional interpretations of WMD and CBRN. However, it also has been shown that these attempts either run into new problems or fail to resolve the original problems. However, the score-based approach of WME that was introduced by Yengst seemed to be the most promising attempt to cope with the problems that the term “WMD” poses with regard to terrorism. Hence, elements of Yengst’s methodology will form the basis for my own proposal. In the following section, a dynamic rating system to identify the most dangerous terrorist weapons will be introduced.

Obviously, the term “dangerous” in this context is vague and unhelpful, at least at first glance. However, on my account dangerous will be cashed in terms of the broader concept of risk. Thus, a dangerous terrorist weapon is a weapon that poses the greatest risk to society. As several researchers already pointed out, risk is a two-dimensional term that refers both to the harmful impact as well as the probability of that impact (See e.g. Forest [29]). Thus, in the cases of terrorist weapons the risk would be calculated by recourse to, firstly, the factors restricting the terrorist’s decision to use a weapon and, secondly, the possible impact (or effect) that this weapon would have if used by terrorists. As already seen above, Yengst’s criteria for defining WMEs are primarily aimed at one dimension of the risk that a terrorist weapon poses, namely the impact (or effect) of the weapon. However, to properly analyze this risk, both dimensions, impact and probability, are needed. Hence, the rating system in this section will not only include some of the criteria Yengst uses to assess the impact of a certain weapon but will also identify factors on the probability axis—in particular, factors that restrict the weapon choices of terrorists.

Assessing the likelihood with which a weapon might be used by terrorists is a highly complex endeavor. Terrorist groups and lone operators are agents with a wide variety of motives (both rational and irrational) who are also interested in disguising their decision-making and in deceiving researchers and investigators. Thus, a quantitative and standardized estimation of the probabilities of the use of certain weapons by terrorists is, in general, challenging. However, researchers like Gary Ackerman, Adam Dolnik, Brian Jackson, and others have identified and discussed several criteria that might influence the decision making of a terrorist group to use a specific weapon for an attack [1, 17, 21, 41]. Based on these criteria, it might be possible to give an indication as to how likely it is that a terrorist group might be successful in acquiring and using a certain weapon for an attack.

First of all, however, it is necessary to further refine the criteria to be used in assessing the impact or effect of a certain weapon in the hands of terrorists. One can, at least, identify four major criteria that contribute to the impact of a certain weapon:

## (a) Hard damage

First of all, the most visible impact that a weapon can produce is physical damage. This damage includes destruction of, and physical damage to, buildings or other structures as well as the physical harming or killing of persons and animals. However, while damage to buildings and persons can be easily characterized as physical damage, that might not be as easy with other forms of damage, such as the damage created by a cyber-attack. Since no kinetic force is used to conduct these attacks, but rather digital means such as software, it might be difficult to call the damage inflicted by a cyber-attack hard damage.<sup>4</sup> However, I argue that, depending on the chain of consequences caused by a cyber-attack, one should characterize its damage as hard damage even if the direct damage caused by the attack might not be physical. This argument holds especially for those cyber-attacks directed at critical infrastructure. In most of these cases, the software is not the weapon itself but rather the means to turn the critical infrastructure into some sort of second-degree weapon that, via being destroyed or damaged, does physical harm to persons or damage to buildings.

In addition to physical harm or damage resulting from an attack, international organizations such as the ICRC stress that other specific harms that are not of a physical nature can have devastating destructive effects on civilian life as well. With regard to these harms, the ICRC counts (amongst others) (1) mental harm as well as (2) economic loss and displacement, as potentially having such a destructive effect [38], pp. 35–37 and 41–43.

- (1) Mental harm as one possible source of damage in the aftermath of an attack is, according to the ICRC, implicitly mentioned in international humanitarian law since it forbids “(...) acts or threats of violence the primary purpose of which is to spread terror among the civilian population” (opp. cit. [38], p. 33). In this quote, “terror” refers to severe mental harm in the form of horror, psychological trauma, and post-traumatic stress.

Two important examples of such psychological reactions to terrorist attacks are anxiety and moral outrage. A terrorist attack with an advanced weapon technology or CBRN device has the potential to inflict widespread anxiety in society [2], p. 24; [3, 33, 52]. For example, public fear of possible contamination caused by improvised radiological or chemical weapons would be instances in which weapons inflict a massive degree of anxiety [44, 52, 66].

Moral outrage can be understood as the anger and horror at the severe violation of a moral standard [7], p. 155. Hence, the complex emotion of moral outrage does not only include anxiety and horror, but primary anger and disgust that can manifest in demonstrations, public condemnations of attacks or calls for justice on a collective level [43]. Arguably, those attacks performed with unconventional and globally ostracized weapons (such as chemical or biological agents) have the potential to cause a larger degree of moral outrage than, for example, an attack with a knife or gun.

---

<sup>4</sup> See Adam Henschke’s chapter in this book on cyberterrorism and the internet of things for more on this discussion.

While a certain degree of anxiety and moral outrage seems, at first glance, a proportionate reaction to an attack, and is in many cases only a temporary condition with minor influence on the impact of an attack, both anxiety and moral outrage can, depending on the nature of the attack, result in political militancy or in calls for (disproportionate) retaliation [33, 34, 58]. One effect of this could be the erosion of trust in security institutions. For example, a successful attack with an impactful weapon might harm the reputation of intelligence institutions, law enforcement, and the military since it may result in the public ceasing to trust them and their ability to keep society safe [50], p. 214; [65], p. 11.<sup>5</sup>

(2) Economic loss and (at least temporary) displacement could add to the impact of a terrorist attack. Particularly, those attacks that involve weapon technologies capable of causing contamination of a certain area potentially cause significant economic damage [48] by means of rendering a certain area (e.g., a business or shopping buildings or streets) unusable for a long period of time. It is noteworthy that not only a de facto-contamination of a certain area would cause economic damage, but also the public fear of contamination in the aftermath of, for example, a radiological attack that was, in fact, not capable of causing any health-damaging contamination (See Khripunov [44]).

(c) Length of the attack

Not only the damage caused by an attack with a certain weapon but also the attack itself can tell a lot about the impact of said weapon. One important factor is the length of the attack in terms of the duration of use of this weapon during an attack. For example, a knife is a weapon that demands multiple uses over a long duration to create significant physical damage (i.e., to harm many people). In contrast, an IED is able to create large scale damage in a one-off use. Other than in case of a knife attack, security forces responding to an IED attack do not have any chance to interrupt or stop the attack as it happens. Hence, a weapon that creates significant damage in a very short time can be characterized as especially impactful.

(d) Uncertainty of consequences

Contrary to Yengst's approach, it may be very hard (if not impossible) to properly anticipate the damage a certain weapon will do in terms of physical, economic, and psychological damage. However, arguably the impact of a certain weapon should be considered especially high if one is unable to anticipate the consequences resulting from the use of it. This uncertainty associated with a particular weapon extrapolates its psychological damage by means of spreading large-scale fear in public. For example, the severity of the consequences from the use of pathogens as terrorist weapons is a matter of controversy among experts, yet the public believes the effects of biological weapons to be catastrophic [42, 52], pp. 6–7, [62]. The town of Salisbury was

---

<sup>5</sup> Please note that several empirical studies found that the aftermath of a terrorist attack can also have the potential to temporarily increase trust in the Government and in other members of society in general. This effect is known as the rally effect. However, recent studies showed that this effect is only a short term effect in the immediate aftermath of an attack [20, 32, 65].

extensively contaminated with the most deadly chemical agent ever produced (Novichok), yet only three people were wounded as a result of this attack [27]. However, the uncertainty concerning the effects of terrorists using biological weapons makes these weapons especially effective in terms of causing psychological and other forms of soft damage. With regard to counter-measures against these weapons, security agencies often refer to the precautionary principle as a guiding approach (General discussion concerning this principle in Grunwald [35], Roeser et al. [57]).

However, the uncertainty attached to these weapons is a problem not only for the counter-terrorism authorities but also for the individual who uses them. First of all, as is the case for the authorities, the perpetrator faces a high degree of uncertainty with regard to the extent of the impact a certain, advanced weapon would have. For example, the release of a fatal virus in a shopping center might have a tremendous impact, yet the fragile nature of viruses as well as environmental conditions and other factors might diminish said impact dramatically. Secondly, the perpetrator of such an attack faces uncertainty with regard to her own security when using certain weapon types. For example, in the example above the perpetrator might very well fall victim to her own weapon during the attack against the shopping mall. This dual uncertainty makes it almost impossible to use said weapons in a controlled and discriminate manner. This uncontrollability makes these weapons even more dangerous and, hence, increases their potential impact.

So far, these four criteria only give information about what could happen *if* terrorists would acquire and use a certain weapon technology. However, to properly analyze the risk certain weapon types are posing, it is also necessary to consider the factors that increase or decrease the probability that terrorists might acquire and use a certain weapon. In addition to the criterion of high impact of a weapon, researchers have shown that terrorists might also consider the following criteria in choosing their weapons:

(a) Availability

The probability that a certain weapon will be used by terrorists can be seen as high if the materials that are necessary to assemble said weapon are openly available or can be acquired with little restrictions. Furthermore, the financial means that are necessary to acquire and assemble a particular weapon are part of the decision-making process of a terrorist group in their choice of weapons. The more affordable a weapon is, the more likely it will be acquired by small cells and lone operators [1], pp. 14, 76–82, 90, Fig. 4.1; [17], Table 2.1; [16], pp. 48–57, [21], p. 19, [31], pp. 1–13, [30], pp. 269–282, [40], pp. 198–201.

(b) Required expertise

Expertise plays a crucial role in the acquisition and use of weapons by terrorists. Some weapon types require extensive and specialized expertise to be used successfully, while others do not require deep knowledge of any kind. Here, the pre-existing expertise as well as the knowledge resources (i.e., personnel, network, safe spaces for testing) of a terrorist group deeply influence what kind of weapon will be chosen for an attack [1], pp. 14, 83, 87–88, [17], Table 2.1, [31], pp. 1–13, [30], pp. 269–282.

## (c) Operational space needed

One particularly important factor determining the expertise that is needed to successfully use a certain weapon is the sophistication of the delivery system for such a weapon<sup>6</sup>. A weapon with a specialized, complex delivery system might create a large impact, but might require a large amount of resources and considerable specialized expertise. Some weapon technologies need extensive space and specialized facilities if they are to be used in an attack. For example, the construction of an improvised nuclear device (IND) requires, at least, a laboratory with specialized equipment and facilities to store raw materials, precursors, and other materials. In a similar fashion, the handling of pathogens such as *Yersinia pestis* (the bacteria that causes the plague) demands laboratory conditions with suitable safety standards to avoid accidental infection. Yet, a simple IED might be manufactured in an apartment in an urban area without risking detection.

The operational space that is needed to manufacture a certain weapon type influences the weapon choices of terrorists in, at least, two ways: first of all, a large operational space such as an industrial complex, a laboratory or a remote facility requires very considerable financial resources. Secondly, a large operational space increases the risk of detection by security agencies. Potential terrorists would have to sign documents and create cover stories in order to get access to a laboratory facility. These procedures make them and their plot vulnerable to being exposed and interrupted [4, 12, 16, 23, 28, 47].

## (d) Tactical, strategical, and ideological advantage

Last but not least, the use of a particular weapon has to have a clear tactical, strategic or ideological advantage over other weapons. Some terrorist groups have a strategy of toppling a regime by targeting specific persons and institutions, while others prefer to spread fear with mass-casualty attacks. Hence, the strategy and, consequently, the preferred tactics of a group determine the weapon choice of a terrorist group as well [1], pp. 13, 72, 99, [17], Table 2.1, [21], pp. 13–21, [41], p. 15.

However, not only tactics and strategy but also the underlining ideology of the group plays a crucial role here [1], pp. 12, 73, 83, [17], p. 44, [21], p. 70f, [22]. For example, a Marxist-Leninist terrorist group that mainly targets political figure-heads might not be as interested in indiscriminate biological agents as an apocalyptic religious group that attempts to kill all “infidels”.

It is important to note that all of these weapon choice criteria cannot be understood as general rules for terrorist decision-making. Rather, they should be seen as indicators for weapon choices that are highly dependent on specific ideologies, organizational structures and capabilities of terrorist groups [1, 17, 21, 41, 45]. For example, the weapon choice pattern of so-called Islamic State of Iraq and the Levant inspired lone operators in Western Europe might be completely different from the weapon choice pattern of the Revolutionary Armed Forces of Colombia (FARC) in Colombia. Hence, to accurately assess the risk that a particular weapon poses, one

---

<sup>6</sup> The author expresses his gratitude to Michael L. Gross for raising this point [15].

has to specify this risk by means of attaching it to a certain terrorist branch (e.g., Islamist cells or right-wing lone operators) and a region (e.g., Western Europe).

Furthermore, the assessment of the impact that a certain weapon might have cannot necessarily be generalized. To properly assess the impact of a weapon, it is important not only to avoid general weapon categorizations, such as CBRN or CBRNE, one should also avoid generalizations of weapon types such as “chemical weapon” or “explosive”. Rather, one should attempt to focus on the nature and amounts of ingredients that a particular weapon consists of to arrive at a specific scenario that can be coupled with the specified weapon choice patterns of a particular group in a particular region. For example, one could assess the impact of a medium-sized improvised chemical device consisting of phosphine and estimate whether the choice patterns of a small terrorist cell in a Western democracy would be in favor of this weapon.

## 7 Conclusion

This chapter has shown that the categorization of weapon technologies using concepts like WMD runs into severe problems when applied to the phenomenon of terrorism. Hence, it was proposed to abolish the static approach that lists weapon categories with regard to the terrorist threat and, instead, to introduce a dynamic rating system to assess the risk that specific weapons pose in the hands of particular terrorist groups.

Yet, to what degree is this rating-based approach superior to the above discussed CBRNE interpretation of WMD that is (at least to some degree) currently being used in counter-terrorism practice? First of all, from a conceptual perspective, the rating approach has the advantage of giving a more detailed overview of the risk that a certain weapon type poses in the hands of a given terrorist group. Not only physical impact and casualty numbers but also soft damage and the handling of the weapon technology as well as its availability and ease of use are included in this overview. Secondly, the rating approach does not include or exclude a fixed set of weapon types. Therefore, this approach can be used to determine the risk of a wide variety of weapons that might be used by terrorists in the future. Thirdly, the approach to use a rating system for these weapons with regard to terrorism does not conflict with the existing definition of WMD in international legislation. After all, a nuclear weapon can be both a WMD according to international law and the most impactful (yet least available) terrorist weapon on the scale.

Additionally, from the point of view of practitioners and counter-terrorism institutions, the more detailed account of the presumed impact of a certain weapon in the hands of terrorists could be used to allocate resources more efficiently on particular weapon types that pose the greatest risk. After all, the counter-measures against the acquisition of an off-the-shelf nuclear weapon might be radically different from the counter-measures necessary to prevent an attack with the above-described improvised phosphine device or a crude RDD. While the first one requires international efforts of non-proliferation and the enforcement of international treaties, the latter

one involves counter-measures such as educating and cooperating with hardware store employees or companies that produce pesticides in Western democracies on a local level. Hence, the introduced weapon rating system enables counter-terrorism institutions to group certain weapon types together dynamically and allocate specific groups of counter-measures necessary to prevent attacks using said weapons.

Finally, the above introduced rating system can help to resolve some ethical issues that arise from the use of concept (and mis-conceptualisations) of WMD in counterterrorism practice. Using the suggested rating system would prevent practitioners and other stakeholders like the press from misinterpreting terrorist plots with small amounts of toxic or radiological substances as WMD events. A more complex understanding of the impact of terrorist attacks with these substances can help to prevent disproportionate responses to threats by police forces as well as exaggerated and fear-inducing reporting by the news media in the aftermath of an attack or foiled plot.

## References

1. Ackerman G (2014) “More bang for the buck”: examining the determinants of terrorist adoption of new weapons technologies. King’s College London (University of London)
2. Ackerman G, Jacome M (2018) WMD terrorism. *PRISM* 7(3):22–37
3. Ackerman GA, Pinson LE (2014) An army of one: assessing CBRN pursuit and use by lone wolves and autonomous cells. *Terrorism Polit Violence* 26(1):226–245. <https://doi.org/10.1080/09546553.2014.849945>
4. Ackerman GA, Pereira R (2014) Jihadists and WMD: a re-evaluation of the future threat. *CBRNe world*, 27–34
5. Asal VH, Ackerman GA, Rethemeyer RK (2012) Connections can be toxic: terrorist organizational factors and the pursuit of CBRN weapons. *Stud Conflict Terrorism* 35(3):229–254. <https://doi.org/10.1080/1057610X.2012.648156>
6. BBC (2006) Blair defends police terror raid. BBC News. [http://news.bbc.co.uk/2/hi/uk\\_news/politics/5053618.stm](http://news.bbc.co.uk/2/hi/uk_news/politics/5053618.stm)
7. Batson CD, Chao MC, Givens JM (2009) Pursuing moral outrage: anger at torture. *J. Exp. Soc. Psychol.* 45(1):155–160
8. Bentley M (2012) The long goodbye: beyond an essentialist construction of WMD. *Contemp Secur Policy* 33(2):384–406. <https://doi.org/10.1080/13523260.2012.693804>
9. Bentley M (2014) Weapons of mass destruction and US foreign policy the strategic use of a concept. Taylor and Francis
10. Binder MK, Ackerman GA (2019) Pick Your POICN: introducing the profiles of incidents involving CBRN and non-state actors (POICN) database. *Stud Confl Terrorism* 24:1–25. <https://doi.org/10.1080/1057610X.2019.1577541>
11. Brown KE (2010) Contesting the securitization of British muslims. *Interventions* 12(2):171–182
12. Bunker RJ (2000) Weapons of mass disruption and terrorism. *Terrorism Polit Violence* 12(1):37–46. <https://doi.org/10.1080/09546550008427548>
13. Carus WS (2012) Defining weapons of mass destruction. DTIC Document
14. Carus WS (2017) A short history of biological warfare: from pre-history to the 21st century
15. Caves Jr JP, Carus WS (2014) The future of weapons of mass destruction: their nature and role in 2030. DTIC Document



16. Cragin, K., Daly, S. A., Everingham, S. S., Hoube, J., Kilburn, M. R., & Marcum, C. Y. (2004). *The dynamic terrorist threat: An assessment of group motivations and capabilities in a changing world*. Rand Corporation.
17. Cragin K (2007) Sharing the dragon's teeth: terrorist groups and the exchange of new technologies, vol 485. Rand Corporation
18. Danzig R, Sageman M, Leighton T, Hough L, Yuki H, Kotani R, Hosford ZM (2011) Aum Shinrikyo. insights into how terrorists develop biological and chemical weapons
19. Deutsche W (2018) Die neue Sorge vor der Bio-Bombe. Dw.De. <https://www.dw.com/de/die-neue-sorge-vor-der-biobombe/a-44326086>
20. Dinesen PT, Jæger MM (2013) The effect of terror on institutional trust: new evidence from the 3/11 Madrid terrorist attack. *Polit Psychol* 34(6):917–926
21. Dolnik A (2007) Understanding terrorist innovation: technology, tactics and global trends. Routledge
22. Drake CJM (1998) The role of ideology in terrorists' target selection. *Terrorism Polit Violence* 10(2):53–85
23. Dunn LA, DeMarce A, Givner-Forbes R, Grosiak A, Kovner M, Lukasiak SJ, Moran N, Skypek T, Yengst W, Perry JL (2008) Next generation weapons of mass destruction and weapons of mass effects terrorism, pp. [2–5] 4–5
24. Early BR, Martin EG, Nussbaum B, Deloughery K (2017) Should conventional terrorist bombings be considered weapons of mass destruction terrorism? *Dyn Asymmetric Conflict* 10(1):54–73. <https://doi.org/10.1080/17467586.2017.1349327>
25. Enemark C (2011) Farewell to WMD: the language and science of mass destruction. *Contemp Secur Policy* 32(2):382–400. <https://doi.org/10.1080/13523260.2011.590362>
26. Enemark C (2012) The unfinished business of abandoning WMD: a reply to Bentley. *Contemp Secur Policy* 33(2):407–412. <https://doi.org/10.1080/13523260.2012.693806>
27. Faulconbridge G, Holden M (2018) Explainer: the poisoning of former Russian double agent Sergei Skripal. Reuters.Com
28. Flade F (2016) The Islamic state threat to Germany: evidence from the investigations. *CTC Sentinel* 9(7):11–14
29. Forest JFF (2012) Framework for analyzing the future threat of WMD terrorism. *J Strateg Secur* 5(4):51
30. Forest JFF (2008) Knowledge transfer and shared learning among armed groups. In: Norwitz JH (ed) *Armed groups: studies in national security, counterterrorism, and counterinsurgency*. Dept. of the Navy, pp 269–289
31. Forest JF (2006) *Teaching terror: strategic and tactical learning in the terrorist world*. Rowman & Littlefield Publishers
32. Geys B, Qari S (2017) Will you still trust me tomorrow? The causal effect of terrorism on social trust. *Public Choice* 173(3):289–305
33. Gross ML, Canetti D, Vashdi DR (2016) The psychological effects of cyber terrorism. *Bull At Sci* 72(5):1–8
34. Gross ML, Canetti D, Vashdi DR (2017) Cyberterrorism: its effects on psychological well-being, public confidence and political attitudes. *J Cybersecur* 3(1):49–58
35. Grunwald A (2008) Nanoparticles: risk management and the precautionary principle. In: Jotterand F (ed) *Emerging conceptual, ethical and policy issues in bionanotechnology*. Springer, Netherlands, pp 85–102. [https://doi.org/10.1007/978-1-4020-8649-6\\_6](https://doi.org/10.1007/978-1-4020-8649-6_6)
36. House CN (2016) The chemical, biological, radiological, and nuclear terrorism threat from the Islamic state. *Mil Rev* 96(5):68–75
37. Hummel S (2016) The Islamic state and WMD: assessing the future threat. *CTC Sentinel* 9(13):18–21
38. ICRC (2016) Principle of proportionality in the rules governing the conduct of hostilities under international humanitarian law
39. Ivanova K, Sandler T (2007) CBRN attack perpetrators: an empirical study. *Foreign Policy Anal* 3(4):273–294. <https://doi.org/10.1111/j.1743-8594.2007.00051.x>

40. Jackson BA (2001) Technology acquisition by terrorist groups: threat assessment informed by lessons from private sector technology adoption. *Stud Conflict Terrorism* 24(3):183–213
41. Jackson BA, Frelinger DR (2008) Rifling through the terrorists' arsenal: exploring groups' weapon choices and technology strategies. *Stud Confl Terrorism* 31(7):583–604
42. James LC, Oroszi TL (2015) *Weapons of mass psychological destruction and the people who use them*. Praeger
43. Johansen ML, Sandrup T, Weiss N (2018) Introduction: the generative power of political emotions. *Conflict Soc* 4(1):1–8
44. Khripunov I (2006) The social and psychological impact of radiological terrorism. *Nonproliferation Rev* 13(2):275–316
45. Koehler-Derrick G, Milton DJ (2017) Choose your weapon: the impact of strategic considerations and resource constraints on terrorist group weapon selection. *Terrorism Polit Violence* 31:1–20. <https://doi.org/10.1080/09546553.2017.1293533>
46. LaFree G, Dugan L, Miller E (2014) *Putting terrorism in context: lessons from the global terrorism database*. Routledge
47. Lakoff A (2007) Preparing for the next emergency. *Publ Cult* 19(2):247
48. Lemyre L, Clément M, Corneil W, Craig L, Boutette P, Tyshenko M, Karyakina N, Clarke R, Krewski D (2005) A psychosocial risk assessment and management framework to enhance response to CBRN terrorism threats and attacks. *Biosecur Bioterror* 3(4):316–330
49. Madeira JL (2012) Killing McVeigh: the death penalty and the myth of closure. NYU Press
50. Meyer B (2004) *Fighting terrorism—a narrow path between saving security and losing liberty*. Globalization, Armed Conflicts and Security
51. Organisation for the Prohibition of Chemical Weapons (1992) *Convention on the prohibition of the development, production, stockpiling and use of chemical weapons and on their destruction*
52. Palmer I (2004) The psychological dimension of chemical, biological, radiological and nuclear (CBRN) terrorism. *J R Army Med Corps* 150(1):3–9
53. Parachini J (2003) Putting WMD terrorism into perspective. *Wash Q* 26(4):37–50. <https://doi.org/10.1162/016366003322387091>
54. Pichtel J (2011) *Terrorism and WMDs. Awareness and response*. CRC Press
55. Pillar PR (2006) Intelligence, policy, and the war in Iraq. *Foreign Affairs*, pp 15–27
56. Rheinische Post (2019) Düsseldorf: Rizin-Anschlagspläne waren weit fortgeschritten. Rheinische Post. [https://rp-online.de/nrw/staedte/koeln/duesseldorf-rizin-anschlagsplaene-waren-weit-fortgeschritten\\_aid-39751823](https://rp-online.de/nrw/staedte/koeln/duesseldorf-rizin-anschlagsplaene-waren-weit-fortgeschritten_aid-39751823)
57. Roeser S, Hillerbrand R, Sandin P, Peterson M (eds) (2012) *Handbook of risk theory: epistemology, decision theory, ethics, and social implications of risk*, vol 1. Springer Science & Business Media
58. Shandler R, Gross ML, Backhaus S, Canetti D (2021) Cyber terrorism and public support for retaliation—a multi-country survey experiment. *Brit J Polit Sci* 1–19
59. Spilcker A (2018) Köln: Tunesier festgenommen: Wollte offenbar mit Giftanschlag Ungläubige töten. Focus. [https://www.focus.de/politik/deutschland/koeln-er-wollte-offenbar-mit-einem-giftanschlag-die-unglaeubigen-toeten\\_id\\_9092406.html](https://www.focus.de/politik/deutschland/koeln-er-wollte-offenbar-mit-einem-giftanschlag-die-unglaeubigen-toeten_id_9092406.html)
60. (START), N. C. for the S. of T. and R. to T. (2016) *Global terrorism database* [Data file].
61. Staudenmaier R (2018) German police carry out more raids in Cologne after charging man with making biological weapon. Deutsche Welle. <https://www.dw.com/en/german-police-carry-out-more-raids-in-cologne-after-charging-man-with-making-biological-weapon/a-44236311>
62. Sullivan GR, Bongar B (2007) Psychological consequences of actual or threatened CBRNE terrorism. *Psychology of terrorism*, pp. 153–163
63. The Times of Israel (2015) Israel tests “dirty bombs,” finds they pose no substantial danger. The Times of Israel. <http://www.timesofisrael.com/israeli-tests-find-dirty-bombs-pose-no-substantial-danger/>
64. United Nation Office of Disarmament Affairs (1975) *The convention on the prohibition of the development, production and stockpiling of bacteriological (biological) and toxin weapons and on their destruction*. un.org

65. Van Der Does R, Kantorowicz J, Kuipers S, Liem M (2019) Does terrorism dominate citizens' hearts or minds? The Relationship between Fear of Terrorism and Trust in Government. *Terrorism and Political Violence* 1–19
66. Wessely S (2005) Don't panic! Short and long term psychological reactions to the new terrorism: the role of information and the authorities. *J Ment Health* 14(1):1–6. <https://doi.org/10.1080/09638230500048099>
67. Westdeutscher Rundfunk (2018) Rizin-Fund in Köln: Tunesier mischte Bio-Waffen zusammen. WDR.De. <https://www1.wdr.de/nachrichten/rheinland/koeln-chorweiler-toxische-substanzen-100.html>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Terrorism and the Internet of Things: Cyber-Terrorism as an Emergent Threat



Adam Henschke

**Abstract** In this chapter I present an argument that cyber-terrorism will happen. This argument is premised on the development of a cluster of related technologies that create a direct causal link between the informational realm of cyberspace and the physical realm. These cyber-enabled physical systems fit under the umbrella of the 'Internet of Things' (IoT). While this informational/physical connection is a vitally important part of the claim, a more nuanced analysis reveals five further features are central to the IoT enabling cyber-terrorism. These features are that the IoT is radically *insecure*, that the components of the IoT are *in the world*, that the sheer numbers of IoT devices mean potential attacks can be *intense*, that the IoT will likely be powered by a range of Artificial Intelligence aspects, making it *inscrutable*, and that the IoT is largely *invisible*. Combining these five factors together, the IoT emerges as a threat vector for cyber-terrorism. The point of the chapter is to go beyond recognising that the IoT is a thing in the world and so can enable physical impacts from cyber-attacks, to offer these five factors to say something more specific about just why the IoT can potentially be used for cyber-terrorism. Having outlined how the IoT can be used for cyber-terrorism, I attend to the question of whether such actions are actually terrorism or not. Ultimately, I argue, as the IoT grows in scope and penetration of our physical worlds and behaviours, it means that cyber-terrorism is not a question of if, but when. This, I suggest, has significant ethical implications as these five features of the IoT mean that we ought to be regulating these technologies.

## 1 Cyber Terrorism Has Not Taken Place

In 2013 Thomas Rid published his book *Cyberwar Will Not Take Place* [48]. It has been the topic of considerable attention, with many people offering criticisms on a range of points that he makes [4]. However, despite a world that is facing increasing instability in its geopolitics, and as a range of high-profile information operations

---

A. Henschke (✉)

Philosophy Section, University of Twente, Enschede, The Netherlands

e-mail: [a.henschke@utwente.nl](mailto:a.henschke@utwente.nl)

© The Author(s) 2021

A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,

Advanced Sciences and Technologies for Security Applications,

[https://doi.org/10.1007/978-3-030-90221-6\\_5](https://doi.org/10.1007/978-3-030-90221-6_5)

show the centrality of cyberspace to national security, Rid's titular premise has held out—nothing in the cyber realm has met the criteria to make it an act of war. Stuxnet is instructive here. More than ten years after the event, Stuxnet is still one of the most high-profile cyber-attacks due to it being proof of concept that cyber-attacks can cause physical impacts. Yet, in line with Rid's point, Stuxnet is important because of its uniqueness. There is yet to be another cyber-attack that brings about physical impacts, or at least, one that is publicly known. And definitely nothing that rises to a level that would classify as armed attack. So, despite Rid's arguments facing criticism, his conclusion seems to be holding out.

Looking at terrorist use of the internet, despite their highly sophisticated use of the internet for recruitment, radicalisation, and propaganda [5], even the so-called Islamic State (IS) at their peak did not manage to engage in cyber-terrorism *proper*, see following for what that means. As Julian Droogan and Lise Waldek point out, “in the realms of academia, policy and the media [have] provided many foreboding and even doomsday warnings about the future of cyber-terrorism, which in the main have failed to come to realization” [16]. So-called IS used cyberspace to motivate and guide a range of terrorist acts [5], and as counter-terrorism actors stepped up their actions—including actions in cyberspace—to disrupt larger high profile terrorist activities around the world [7], so-called IS evolved their strategies [29–31] to encourage low technology small group acts of terrorism, using whatever technologies they had at hand—as a spokesperson for so-called IS stated in 2014, “If you are not able to find an IED or a bullet, then single out the disbelieving American, Frenchman, or any of their allies. Smash his head with a rock, or slaughter him with a knife, or run him over with your car, or throw him down from a high place, or choke him, or poison him” [44]. Yet, despite their evolution toward small scale ongoing acts of domestic terrorism, even so-called IS did not mount any successful cyber-terrorism acts. To be clear, so-called IS did use the internet for cyber-attacks [2]. However, insofar as terrorism necessarily involves physical violence, or the credible threat of physical violence, they did not engage in cyber-terrorism.

This turns us to a definition of cyber-terrorism. Terrorism is a complex action that relies on two targets of attack [12, 46]. First is the attack itself. In most accounts of terrorism, the terrorist action uses physical violence to attack people [46] or perhaps their property [12]. The second target is political and social leaders and wider community. It is not simply “the organized use of violence to attack non-combatants (‘innocents’ in a special sense) or their property” but organized violence “for political purposes” ([12], 5). The intent of the terrorist attack is not violence for the sake of violence but that in response to this attack, people's behaviour changes. Ideally, the targeted people and/or their political representatives change some law, policy, practice or behaviour in line with the terrorist's ends.

The issue here is that, to date, terrorist use of the internet has not included efforts where the internet has been used *directly* to bring about *physical violence*. This a vital distinction—if we understand cyber-terrorism to be simply about use of the internet to spread fear or bring about political changes, then we are talking about propaganda or information operations. And while these are important issues, and play a big role in modern international terrorism, I suggest that this is not cyber-terrorism. Contrast

a message posted online that says “we are going to harm you”, with a terrorist action that involved hijacking autonomous vehicles and using the hacked vehicles in coordinated vehicular attacks against pedestrians. Subsequent to the attack, the terrorists broadcast a message that says “we were behind these attacks. And if you don’t follow our demands, we will continue to harm you.” The first example was merely the use of the internet to communicate a threat. The second is the use of the internet to bring about physical violence, coupled with a larger socio-political agenda. While we have seen many instances of the first example, to date we have not seen any instances of the second.

The reason that neither cyber war nor cyber-terrorism have happened, is due in part to the limited capacity for cyberspace to have direct causal impacts in the physical world. The core of Rid’s argument turns out to be true in the world. “[M]ost cyber-attacks are not violent and cannot sensibly be understood as a form of violent action” [13, 48]. The original ‘Tallinn Manual’ holds such a view, exemplified by its Rule 11: “A cyber operation constitutes a use of force when its scale and effects are comparable to non-cyber operations rising to the level of a use of force” [52]. Something like Stuxnet is an aberration; very few cyber-attacks do have the direct impact on the physical realm to count as physically violent. And, in line with the Tallinn Manual’s reasoning, no cyber-attacks have risen to a level comparable to that of physical use of force that would constitute a ‘just cause’ for war. So, descriptively, cyber-attacks simply have not had the physical impacts to be considered war or terrorism. In line with the definition above, I am using cyber-terrorism here to mean something like the use or exploitation of the internet to bring about an act of physical violence directed against non-combatants or innocents, to achieve some secondary ideological, religious or political purpose. Importantly, as will be discussed toward the end of this paper, these acts have to be high profile; they need wide coverage or publicity to ultimately be considered successful. Thus in this description, neither so-called IS nor any other modern terrorist group has used cyberspace to engaged in acts of physical violence to achieve these secondary ends.

However, this is not a permanent fact about cyber-attacks. Looking closer at Rid’s reasoning will help explain why. “Code doesn’t have its own force or energy. Instead, any cyber-attack... has to utilize the force or energy that is embedded in the targeted system or created by it... Computer code can only directly affect computer-controlled machines, not humans” [13, 48]. On Rid’s account, something like a malicious computer virus is something composed of computer code and can only act upon other computer code. A computer virus is importantly different from a biological virus [48, 13–14]. The biological virus directly impacts the host’s body, while the computer virus can only impact other code. According to Rid, code can only act on code.

For the purpose of this chapter, I am accepting Rid’s narrow claim about code-on-code being the only way to conceptualise cyber-attacks and his position that violence is only physical. There is an interesting discussion about narrow/wide definitions of violence,<sup>1</sup> and that if we have a wider view of violence we might rethink what counts

---

<sup>1</sup> For more on different ways conceive of violence see: [13, 15, 19].

as terrorism. Jessica Wolfendale’s chapter in this book touches on some of those issues. My point is that even if we accept Rid’s narrow account about cyber-attacks being code-on-code, and a narrow definition of terrorism that is limited to acts of, or credible threats of, physical violence, the IoT makes cyber-terrorism a meaningful term.

The reason is that the IoT is a cyber-physical system,<sup>2</sup> and so has the capacity for code to have ‘direct’ physical causality. Many elements of the IoT forge a direct link between code and actuators [3]. Actuators are elements which, upon receiving a code-driven command, will bring about some physical change in the world. Think here of a smart car that has remotely activated door locks. Communications between the car owner’s mobile phone and the car mean that the doors will be unlocked as the owner approaches the car. Due to commands from code, the locks move. The code is causing changes in the physical world. Contra Rid, the informational realm is no longer simply code-on-code, it is code-to-world. The IoT exists across, and actively seeks to link the information with the world, the cyber realm and physical realm now have a direct causal connection. As I have argued elsewhere, this combination of cyber and physical realms means we need to consider both in any assessment of the IoT [26]. Moreover, this relation is dyadically causal<sup>3</sup>—the cyber realm influences the physical and the physical influences the information. So, the IoT means that one of Rid’s key premises, that code only acts on code, is no longer correct.

## 2 The IoT: Cyber-Physical Systems That Will Span The Globe

Before going further, we need to clarify what is being referred to when discussing the IoT. In short, this can mean any device or thing in the world that is ‘smart’ and connected with other devices. “‘The IoT’ is a broad, and deliberately vague catch all term to describe a range of integrated technology types that include (1) sensors, ‘things that gather information’, (2) communicators, ‘things that communicate information’ (3) actuators, ‘things that change the physical world’ and (4) AI, things that process information [3, 55–56]. The IoT can include individual devices or components, like a smart TV, a small networked set of devices like a smart home or a large complex system of devices like an autonomous driving system.

---

<sup>2</sup> Groups like the United States National Science Foundation use the term ‘cyber-physical system’. “Cyber-physical systems (CPSs) are transforming the way people interact with engineered systems, just as the Internet transformed the way people interact with information. CPS integrate cyber components (namely, sensing, computation, control, and networking) into physical components (namely, physical objects, infrastructure, and human users), connecting them to the internet and each other” [18, 53, xxix].

<sup>3</sup> A dyadic relation is one that recognises “the idea of mutual causation. There is a particular ‘whole’ which consists in two elements, each of which stands in a causal relation to the other” [25, 267]. For more on this idea of dyadic relations, see [25, 170–173].

The changes brought about by the IoT that make it relevant for a discussion of cyber-terrorism. The IoT enables the informational realm and the physical realm to be causally connected. This occurs through the use of actuators. Actuators are components which allow information, or code, to be translated into changes or impacts in the physical world. It is these actuators that allow information to make these systems cyber-enabled *physical* systems. In addition to the issues of physical safety arising from these actuators, this physicality of the IoT marks it as importantly distinct from the internet. The internet, as we typically understand it, is primarily an informational network. While it exists in, and relies on things in the physical world [33], it is largely constrained to cyberspace. Rid's argument is that because cyber-attacks are code-on-code, their impacts are primarily contained to the cyber realm [48]. The IoT breaks this division. Due to the causal connection between code and actuators, code can now bring about physical impacts.

Moreover, the IoT is expected to be immense. Current estimates "project that there will be more than 41 billion IoT devices by 2027, up from about 8 billion in 2019" [43]. This leads some to predict an investment of 1.7 trillion U.S. dollars by 2020 [32]. The annual investment is now predicted to be \$2.4 trillion by 2027 [43]. Its scale alone will mean that it will bring immense change to our lives. Moreover, the IoT will likely reach into all facets of our lives, the personal in the form of smart homes, the professional in the ways that it will guide working life, the system in how it will affect things like logistics, even the governmental and military.

So, putting these aspects together, we have a scenario where the informational realm and the physical realm are now directly interacting with each other, that may cover the globe and penetrate our personal, professional, social and political lives. Contra Rid, cyberspace is no longer just code-to-code. These elements, I suggest, present terrorists with a capacity to use the internet to cause significant physical violence in order to bring about ideological, religious or political changes. As such, I suggest that cyber-terrorism will take place.

### 3 So What? An Inventory of Features

This chapter could stop at that, but a more nuanced analysis will give us a greater understanding of the particular vulnerabilities of the IoT that make it an ideal novel means for terrorist attacks. In this section, I present an inventory of features that clarify the point that cyber-terrorism will happen. I argue that

- (1) the IoT is radically *insecure*,
- (2) that components of the IoT are *in the world*,
- (3) that the sheer numbers of IoT devices mean potential attacks can be *intense*,
- (4) that the IoT's reliance on AI present further challenges arising from the *inscrutability* of AI, and
- (5) that the IoT is largely *invisible*.



And, in combination, this inventory of features makes the IoT a way for cyber-terrorism to happen.

The IoT is widely acknowledged to be radically *insecure*. This insecurity has led people to describe it as the ‘internet of insecure things’ [10] and ‘the internet of threats’ [41]. In one example of how this insecurity can lead to significant personal risk, one woman was stalked by an ex-partner, who through “simple technology and smartphone apps that allowed him to remotely stop and start her car, control the vehicle’s windows and track her constantly” [54]. A widespread IoT that is integrated into our lives will be like this but many times more powerful and pervasive.

This radical insecurity is brought about by a combination of two aspects of the IoT. As mention above, the IoT is composed of things that are in communication with each other. Not only do many IoT devices have sensors gathering information on the world around them, that information is then communicated. Thus, there will be a wealth of information being shared *between* a range of interconnected components and devices. In an insecure system, that personal information can potentially be accessed by people without the user’s consent. We have seen examples of this with ‘smart toys’, children’s toys with remotely accessible cameras and other sensors present significant security vulnerabilities [11, 24]. A number of smart technology companies have either been shown to be, or publicly admitted to, using cameras and microphones in smart televisions [39, 51]. Devices like Google Home and Amazon’s Alexa [1, 55] allow for remote surveillance in the home. And in perhaps the creepiest example, We-Vibe, a company that produces smart internet connected sex toys, was shown to be gathering user data [47]. We-Vibe were gathering information about how their sex toys were used, the duration and intensity of use, even the temperature of users was gathered and sent back to the company without user knowledge or consent.

This brings us to the radically insecure aspect of the IoT. We-Vibe’s misuse of personal information became apparent when their product was hacked by a group of ‘white hat hackers’.<sup>4</sup> The security on the We-Vibe product was limited at best. This radical insecurity is seen to be pervasive across many IoT devices and products [10, 41]. The basic cyber-security on these things is relatively weak. Second, the passwords that they do have are typically and frequently set to a factory default and then not changed or complex for users to change.

The limited security serves a range of purposes. It makes it easier to install and use the IoT devices. In the ideal scenario, a user buys a device, takes it home, to the office, etc., activates and it merges seamlessly with the communications networks and other relevant devices [9, 14]. However, if one was to have complex security protocols that needed to be run prior to the device coming into operation, this would not only be more time consuming for the user, it would increase the likelihood of connection problems. As anyone who has tried and failed to get Bluetooth devices to pair with each other can attest, connection problems with smart devices can be incredibly frustrating and time consuming. If the connection is not successful, it can either defeat the purpose of purchasing the device and may even render the device

---

<sup>4</sup> White hat hackers are people who hack into devices or information systems to alert owners, users and manufacturers to security vulnerabilities and failings [38].

useless. Further, the limited security keeps costs down. So, ease of use, efficacy of use and costs are values that drive design and security defaults toward lower security features.<sup>5</sup>

Adding to this, malicious agents can access information about factory default passwords online, and so gain access to the information gathered and communicated by the devices. Shodan, for instance, “is a search engine for exploring the Internet and thus finding connected devices. Its main use is to provide a tool for cybersecurity researchers and developers to detect vulnerable Internet-connected devices without scanning them directly” [17]. While this insecurity alone does not necessarily mean that an IoT device could be weaponised by terrorists, the radical insecurity is part of a set of features that make the IoT an ideal target for tech savvy terrorists. Think here of an autonomous vehicle with weak security—should a terrorist discover that this security vulnerability allows for remote control of steering, breaking or accelerating features, the car becomes part of a terrorist attack. And if a series of attacks occurred, not only would that likely cause significant damage to trust in autonomous vehicle systems [27] it could neatly fit with certain terrorist’s second order aims, a point I return to at the end of this chapter.

The second feature of the IoT that makes it a potential target is that these devices are *in the world*. We have already touched on the way that the code-to-world aspect of the IoT means that it could allow for terrorists to bring about physical violence. This is because the IoT is not constrained to the informational realm. It is in the world, and so—depending on the particular devices—can allow for a cyber-attack to bring about physical violence. Think again of an autonomous vehicle being taken over by terrorists. The deliberate use of cars and trucks in terrorist attacks around the world [42] show how vehicles are an increasing weapon of choice for terrorists. With autonomous vehicles this could be done remotely. Of course, autonomous vehicles with such security flaws would likely not be allowed on the road. My point here is that the elements of IoT components that are in the world means that certain IoT devices, like cars, can potentially be used for physical violence.

We can also think about the security challenges posed by the IoT being in world in a different way. Think here that IoT users are primarily civilians, non-combatants or non-security actors. This means that those users are likely not going to have concerns about terrorists using the IoT against them. However, the familiarity with IoT devices can breed lax security practices. For instance, consider security sector actors, like those in military, intelligence, diplomatic or policing roles using IoT devices with a civilian mindset. The point here is that in a world of ‘bring your own device’, those from the security sector need to be extra careful with IoT devices. Consider here the example of the Strava fitness tracking app. Strava was an IoT device in which people’s exercise habits were monitored and shared to publicly accessible social media. A junior university researcher was interested in this publicly accessible information and used it to identify US military and spy bases.

Strava, a fitness-tracking app, is revealing potentially sensitive information about military bases and supply routes via its global heatmap website. The data map shows 1 billion activities

---

<sup>5</sup> I have argued this point in more detail in the design of autonomous vehicles.

and 3 trillion points of latitude and longitude from “Strava’s global network of athletes”, according to the American company. On the weekend, 20-year-old Australian university student Nathan Ruser noticed the map showed the locations and running routines of military personnel at bases in the Middle East and other conflict zones... While security analysts often use satellite imagery to study military installations, Mr Ruser said the Strava data added an additional, possibly dangerous layer of information. Using satellite imagery, you can see base buildings, for example. But on the heatmap, you can see which buildings are most used, or the jogging routes of soldiers [8].

The point here is twofold. First, IoT devices can provide security sensitive information if user behaviour considers these devices with a civilian mindset. That is, because we are familiar with them in a non-security context, we can easily overlook the security threats that they pose. Second, as these devices are in the world, upon analysis they can provide interested parties with useful information about user habits in the world. This derived information can then pose security risks. Whether it is habits of security personnel on military bases, or more general civilian habits like driving patterns, such information derived from the physical presence of these devices can be very useful to terrorists and other malicious actors. I have written elsewhere how the collection, aggregation and analysis of innocuous information can reveal virtual identities of people [25]. The IoT will only add to this capacity to gain increasingly revealing and powerful information about people, which then has significant security implications.

This alone would not seem too relevant to cyber-terrorism. However, when you combine the radical insecurity with the fact that there are billions of IoT devices in the world, you have the potential for *intense activity*. The point here is that malicious actors like terrorists can exploit IoT’s numbers for cyber-attacks. Consider that there have been cyber-attacks that have used ‘smart devices’ like smart fridges with poor security for DDOS attacks [37]. As mentioned, by 2027 some estimate that there will be more than 40 billion IoT devices in the world [43]. The sheer numbers of IoT devices mean that it can act as a force multiplier. As the DDOS examples show, the IoT can be harnessed for other cyber-attacks. Similarly, the number of IoT devices mean that the effects of an IoT attack can potentially be disastrous. Consider here if a smart house has an unsecured IoT enabled heater. If an attacker was to take over this heater, they could turn the temperature of the house up remotely, which is obviously of minimal concern. However, if this attack took over hundreds of thousands of these heaters during a heat wave, it could bring down regional power supply, potentially increasing the number of vulnerable people like the elderly that can die during the heat wave. Thus, the sheer number of IoT devices in the world mean that critical infrastructure is vulnerable to cyber-attacks.<sup>6</sup>

The point here is that, not only does the IoT allow for code to impact the physical realm, but the sheer number of IoT enabled devices in the physical world mean that physical things can have significant impacts at a higher level than what a pre-IoT cyber-attack could cause; the number of devices vulnerable to attack means that

---

<sup>6</sup> Note that on the definition of cyber-terrorism above, such cyber-attacks would not yet constitute cyber-terrorism. The exploitation of the IoT for physical violence needs to be in service of some secondary ideological, religious or political purpose.

these attacks can be intense. The number of these devices in the world, coupled with their radical insecurity mean that a malicious actor can use the IoT to bring about significant disruption in the cyber realm and that this can then have physical impacts. While this is might still be code-to-code attacks in a narrow sense, it is enabled by the numbers of IoT devices in the world.

AI is likely to be an increasingly important part of the IoT. This is because there will be so many connected devices in the IoT. “To reap the actual benefit of IoT, it has to be intelligent” [45, 1]. Given the sheer numbers of IoT devices, there will be a cluster of parallel IoT systems that require co-ordination. Whether it is the devices in one IoT system, or the integration of different systems, the only way that the more complex IoT systems and integrated systems will be able to operate seamlessly, at speed, without human interaction is through AI [21]. In addition, the vast amounts of information that will be gathered and communicated by these devices will dwarf what the internet is currently producing: One current estimate suggests that IoT devices generate 1 billion GB of data each day [21]. Again, the only way that this can be managed is through AI. The problem with AI is that it can be *inscrutable*.<sup>7</sup>

This, inscrutability I suggest, presents an ideal point of vulnerability for terrorists to exploit. Trust is essential for autonomous vehicle systems to function effectively [27], and I suspect that this claim will hold for many IoT systems. If people do not trust the system, they are either not going to use it, or will not use it to its full effect. However, given the inscrutability of the system, it might be impossible to prove that the decision support systems provided by the AI are safe or reliable. The inscrutability of the AI allows for terrorists to exploit confusion and sew mistrust. On its own, inscrutability is not a major terrorist risk, but couple the AI with the IoT being in the world, allowing for intense activity and its radical insecurity and you have a viable threat vector for terrorist activity.

The final feature that means the IoT is a viable threat vector for cyber-terrorism is that it is *invisible*. This invisibility occurs in a range of layers. The actual components of the IoT are going to be typically invisible—cameras in televisions, microphones in smart watches, locks in car doors. A key technological development enabling the IoT is the miniaturization of its components.<sup>8</sup> The sensors, the communicators, the actuators, these technological components that enable the IoT are all undergoing rapid and substantive miniaturization, allowing them to be integrated into a range of different applications [23, 34]. They can be potentially everywhere in our physical world, and by design, we will literally overlook them. When working as it should, the user should be unaware of the IoT devices and components.

---

<sup>7</sup> Note here that I am agnostic about whether the components of the IoT will be automated or have some form of autonomy. Likewise, I am agnostic whether these systems are just information handling devices or if they come closer to proper intelligent systems. The point of this section remains the same. Nothing for my point relies on the IoT systems being properly autonomous, intelligent, sentient, having moral agency and so on.

<sup>8</sup> For instance, “[r]ecent advancement of miniaturization in manufacturing allows IoT devices to easily be loaded into unmanned drones and vehicles because of miniaturized sizes and light-weight designs” [34, 102]. This miniaturization of sensing devices is predicted to play an increasingly important role in the application of the IoT to healthcare [23, 40].

Further, the people in the IoT are invisible. The invisibility of people in the IoT occurs in a complex interactive set of ways. Remote users can be invisible to other users. The designers, and the choices that they have made in the design and decision features of the IoT's components, are typically invisible to most users. Those people who inhabit roles in the oversight mechanisms are likely to be invisible to users. In a poorly designed systems, the users themselves can often be invisible to designers and oversight bodies.<sup>9</sup> This occurs in part when there are poorly designed features that do not take people into account—an autonomous vehicle that allows for a car to be remotely hacked by a malicious actor for instance has not taken into account the threat posed by some people. In addition, users are often invisible to designers in that it is hard to predict how people will actually use, misuse or hack a piece of technology. Moreover, the complexity of ways that a set of people, using technologies in the real world, in cooperation and competition with each other, makes it very hard if not impossible to predict, design and write laws for every possible combination of use.

Finally, the risks are invisible. While the insecurity, the IoT being in the world, the potential intensity of cyber-attacks and inscrutability alone do not alone necessarily make the IoT a means for cyber-terrorism, *in combination they do*. This is essentially an 'emergent risk'. By this I mean that the combination of these features presents a novel system-level risk that can only be properly understood when looking at the combination of these factors. The combination of the IoT being radically insecure, in the world, intense, and inscrutable is a system level phenomenon that can only be properly explained when seen from the system level. This notion is explained by reference emergence. "Emergence is said to occur when certain properties appear in a system that are novel or unexpected and go beyond the properties of the parts of that system" [35, 277]. We lose explanatory power if we look only at each factor independently. By suggesting that we see the IoT as presenting an emergent risk, we are able to better recognise and understand how it can be used for cyber-terrorism. That is, in combination, we have made the risk visible.

## 4 Will IoT Enabled Cyber-Attacks Be Acts of Terrorism?

I have presented a case that five aspects of the IoT in combination present, not just a risk but, a *terrorist* risk. There are, however, two counter-arguments to engage with before we accept the claim that cyber-terrorism will happen. First, is whether an IoT enabled attack counts as terrorism. Second is whether an IoT enabled act of cyber-terrorism is likely.

For the first counter-argument we return to Rid's scepticism about cyber-attacks being violent. Recall that on Rid's view, a cyber-attack was code acting on code, so

---

<sup>9</sup> In their overview of value sensitive design (VSD), Batya Friedman and David Hendry discuss in great detail the need for effective and ethical design to take in the views, needs, values and practices of a large range of stakeholders, including but not limited to direct and indirect users [20, 35–44].

not physical and therefore not violent. As we have discussed, however, the IoT is a complex set of cyber-enabled physical systems. People are physically vulnerable to the IoT in ways that we are not physically vulnerable to the internet. The five features of the IoT listed above: it is insecure, in the world, intense, inscrutable, and invisible mean that we can reject a position like Rid's—an unsecured set of IoT devices that pose physical risks to people can allow code to act in the world.

However, there is a second aspect to the IoT that perhaps should give us pause to consider an IoT attack, even if it is in the world, is it an act of *terrorism*? While it is plausible to suggest that many IoT enabled acts of terrorism might be limited to physical property and not people, the physical nature of the IoT means that a well thought out terrorist attack puts people at physical risk.<sup>10</sup> The most obvious scenario is that a group is inspired by the way that so-called IS and right wing extremists have started using cars to deliberately drive into groups of people [42]. While such an attack is—arguably—an attack on physical property, the relevant factor is that that physical property is then used to physically harm people. To reiterate a point made above, autonomous vehicle designers take these risks quite seriously so it is hopefully unlikely that such an attack might occur. However, as the IoT becomes more widely dispersed and deeply integrated into our lives and world, the risk of some aspect of it being hacked to cause physical harm to people is something that should not be dismissed. Just as a set of box cutting knives enabled the hijacking of planes on 11 September, all it takes is a creative thinker to exploit some combination of factors in the IoT to engage in an act of cyber-terrorism. And as I have showed with the inventory of five features of the IoT, it presents an attractive target for terrorists. Further to this, as so-called IS showed, modern terrorism is not shy of using either modern information communications technologies or common items like cars to further their terrorist aims. The motivation is there, and the IoT provides the means for cyber-terrorism to occur.

The second counter-argument is scepticism about whether such IoT enabled cyber-terrorism is *likely* to happen. Terrorism is not simply concerned with physical violence against innocent people, but some second order effects. Again, terrorism, it is “the deliberate use of violence, or threat of its use, against innocent people, with the aim of intimidating some other people into a course of action they would not otherwise take” [46, 24]. Essential to any successful act of terrorism is that it brings about the second order political, religious or ideological ends that motivate the group. Or at very least, that the act of terrorism uses physical violence to draw attention to those political, religious or ideological ends. “The success of a terrorist operation depends almost entirely on the amount of publicity it receives...Thus in the final analysis, it is not the magnitude of the terrorist operation that counts but the

---

<sup>10</sup> We have also recently seen that a cyber-attack on a hospital caused a death resulting from disruption to the IT system: “the first known fatality related to ransomware occurred in Duesseldorf, Germany, after an attack caused IT systems to fail and a critically ill patient needing urgent admission died after she had to be taken to another city for treatment.” [6]. This example, however, is an act of cyber-crime, rather than cyber-terrorism, as it lacks the secondary ideological, religious or political purpose necessary to make it an act of terrorism. But it does show how cyber-security can be a matter of life and death.

publicity; and this rule applies not only to single operations but to whole campaigns” [36, 109]. The likelihood of an IoT enabled cyber terrorist act occurring is thus a function of the anticipated publicity that the act will receive.

As Paul Gill et al. note, terrorist groups often display a capacity for “malevolent creativity” [22, 130]. One of the features driving terrorist creativity is the novelty of an attack: “Spontaneous novel acts of violence generate effective surprise within the target audience” [22, 134]. Here, one can only speculate, but the relative novelty of particular IoT systems seems like they are an ideal means of a shocking terrorist act. As these systems are new, they are particularly vulnerable to the fear that results from a terrorist attack. Consider again autonomous vehicles. “If trust is necessary for effective driving, then the background beliefs about *whether* the given technologies and systems are trustworthy will impact how and when people drive. This in turn depends on whether the drivers see other drivers, road users and the system itself as trustworthy. Moreover, once trust is lost it can be very hard to repair” [27, 89]. If an IoT system was to be the subject of a terrorist attack, then it is likely that many users and relevant oversight bodies would either cease using the system or demand significant security changes as they see the overall system as untrustworthy. While in the long term, increased security would ideally reduce the risk of ongoing cyber-terrorism, the fear that a high profile attack would generate and the reduction in use cause by a loss of trust would fit the second order aspect of terrorism. And the fact that changes would be made is evidence of the success of the attack—think here of the security response to air travel following high profile terrorist acts that targeted planes.

The likelihood of such attacks becoming widespread is likely going to be a combination of the amount of public coverage that such attacks generate, and how the publicity around the attacks connects with the larger ideological, religious or political purpose of the terrorist actors. My speculation here is that, at least in the early days of such IoT enabled cyber-terrorism, the attacks will be seen as both novel, and provide some high level of spectacle, thus attracting a lot of publicity.

## 5 Ethics and Responsibilities for IoT Enabled Cyber-Terrorism

To close this discussion, let us put this in the context of ethics. The chief ethical issues here are concerned with responsibility for IoT enabled cyber-terrorism. If, as I have suggested, cyber-terrorism will be enabled by the IoT, then what ought we do about it? The five features described provide us with a way to get some nuanced ascription of responsibility. First, and foremost, the radical insecurity of the IoT needs to be dealt with. This is in part a governmental responsibility—it is national governments who have the capacity to draft and enforce laws that ensure minimum security standards. However, unlike many other areas of counter-terrorism, service



and technology providers also bear some responsibility here. If security vulnerabilities in their products and services that allow for the IoT to enable cyber-terrorism, then it is incumbent upon them to resolve those security failings.

Second, that the IoT is in the world entails a responsibility on governments, technology designers and providers, and consumers to be aware of the risks that their IoT components pose. The point here is that if the products and services that we use in the world provide the infrastructure for cyber-terrorism, then we all have a responsibility to do what we can to mitigate this risk. This would include things like ensuring that our own IoT devices have their security updated and upgraded as necessary. Importantly, such resolutions to the issues of security are not going to happen without recognition of the risks posed by these things in the world.

Third, on the issue of intensity, following the responsibilities for insecurity and the IoT being in the world, if we take the vulnerabilities posed by the IoT seriously, then we should hopefully have significantly reduced the potential for intensity of the cyber-attacks. The responsibility here falls again on governments, technology designers and providers, and consumers.<sup>11</sup>

The inscrutability of AI and its potential role in cyber-terrorism presents a very novel challenge. However, there is a burgeoning literature on the ethical importance of explicability that we can draw from here. “It is rare to see large numbers of ethicists, practitioners, journalists, and policy-makers agree on something that should guide the development of a technology. Yet, with the principle requiring that [AI] be explicable, we have exactly that. Microsoft, Google, the World Economic Forum, the draft AI ethics guidelines for the EU commission, etc. all include a principle for AI that falls under the umbrella of ‘explicability’” [49, 498]. My suggestion here is that explicability, the process by which we reduce inscrutability, needs to be pinned to two parallel principles. When an act of cyber-terrorism appears to have used the IoT we need some processes that can *ensure* that such vulnerabilities are identified and mitigated, and that we can *assure* the public at large that these vulnerabilities are in fact being dealt with.<sup>12</sup> Again, by identifying the feature of inscrutability in IoT enabled cyber-terrorism, we need to find some way of assigning responsibility to governments for oversight, to technology designers and producers to ensure that their products are robust, that can take into account the public facing aspects of the IoT, and its relation to cyber-terrorism.

Finally, to the invisibility of the IoT, we find a further aspect that helps clarify ethical responsibility for such cyber-terrorism. As argued, there are a series of ways that the IoT is invisible to people. The point here is that we generally hold that a person is not to be held responsible for something that they are ignorant of. For instance, if it was my autonomous vehicle that was hacked and used in a terrorist attack, but I was not to know that it presented such a risk, to paraphrase Michael Zimmerman,

---

<sup>11</sup> See also the chapter by Alastair Reed and Adam in this collection for more on this discussion of the responsibility of technology companies around modern terrorism.

<sup>12</sup> In a co-authored article, I have argued elsewhere about the need for insurance and assurance mechanisms with surveillance technologies in liberal democracies, and many of the points there hold here [50].



most would say (and I would again be inclined to agree) that I am not to blame for an act of IoT enabled cyber-terrorism, unless I am to blame for my ignorance.<sup>13</sup> This relation between knowledge, ignorance, and responsibilities is a controversial and contested area. As Zimmerman suggests, we must factor in whether a person is to be blamed for their ignorance, “to say that Perry ought to have known better is to imply that he could have known better—he was free to know better” [57, 413].

However, we can suggest some rules of thumb here—we ought not hold consumers and IoT users responsible for IoT enabled cyber-terrorism if they were reasonably ignorant of the way that their IoT components could be utilised in an act of cyber-terrorism. Given their knowledge of the products and their likely uses, designers and technology producers, however, would have to justify why they were justifiably ignorant that their particular design and products could be used for cyber-terrorism. That is, when thinking of consumers and users, the burden of proof is generally on those seeking to show why the consumer and user ought to be held responsible, while when considering designers and producers, the burden of proof is generally going to be on them to justify why they ought not be held responsible. While each particular instance requires nuance and detail, the rules of thumb are usefully derived from recognition of the invisibility of the IoT. Again, the five features of the IoT give us a way to at least start a nuanced conversation about the ascription of responsibility.

To conclude, in this chapter I mounted an argument that the IoT will enable cyber-terrorism. Given that the IoT is a cyber-physical system, we can reject a Rid style claim that cyber-attacks are only code on code. The causal links between sensors, communicators and actuators mean that a code-based attack can have physical effects. Moreover, I have listed an inventory of five further features that make the IoT a threat vector for terrorism. I showed that the IoT lacks significant security protections making it radically *insecure*. Not only does the IoT pose risks to people’s physical safety in ways that the traditional internet does not, but the fact that its components are *in the world* means it is particularly vulnerable. Add to this the *intensity* of an attack rising from the sheer numbers of IoT devices. Further, as the IoT will require AI to help coordinate components and systems, the decision making may be *inscrutable* which makes for further risk of the second order impacts of cyber-terrorism. Finally, as the IoT is going to be *invisible*, not only will we overlook the components, networks and people involved in its operation, we will also overlook the risks. Combining these five features together, we face an emergent risk from the IoT.

The underpinning factor of how successful IoT enabled cyber-terrorism is, is how resilient the system is to such attacks [27, 28]. By recognising that the IoT is a potential enabler for cyber-terrorism, we are part of the way to reducing the impact of such attacks. The inventory of five features allows us to better understand the risks posed by the IoT. Moreover, by recognising that the IoT is radically insecure, situated in the world, can enable intense outcomes, has elements that are inscrutable, but its risks are invisible, we are better able to understand the ethical responsibility

---

<sup>13</sup> Michael Zimmerman’s original quote is “most would say (and I would again be inclined to agree) that Perry is not to blame for paralyzing Doris, unless he is to blame for his ignorance” [56, 411].

for anticipating and mitigating the risks of cyber-terrorism. So, while I have argued that cyber-terrorism will happen, we do not have to passively allow the terrorists to exploit the vulnerabilities in the IoT. Better design, effective coordinated oversight and a wider public awareness of the risks posed by IoT should help mitigate those risks.

## References

1. AAP (2019) Google listens to user speaker recordings. SBS News. <https://www.sbs.com.au/news/google-listens-to-user-speaker-recordings>
2. Albahar M (2017) Cyber attacks and terrorism: a twenty-first century conundrum. *Sci Eng Ethics Online* First, 1–14. doi: <https://doi.org/10.1007/s11948-016-9864-0>
3. Allhoff F, Henschke A (2018) The Internet of things: foundational ethical issues. *Internet Things* 1–2:55–66. <https://doi.org/10.1016/j.iot.2018.08.005>
4. Allhoff F, Henschke A, Strawser BJ (eds) (2016) *Binary bullets: the ethics of cyberwarfare*. Oxford University Press, Oxford
5. Awan I, Imran A (2017) Cyber-extremism: Isis and the power of social media. *Society* 54(2):138–149. doi: <https://doi.org/10.1007/s12115-017-0114-0>
6. Bajak F (2020) Suspected Ransomware attack Hobbles Major Hospital Chain's U.S. Facilities. PBS News Hour, 29 September. Accessed 29 April 2021. <https://www.pbs.org/newshour/nation/suspected-ransomware-attack-hobbles-major-hospital-chains-u-s-facilities>
7. Blanco JM, Cohen J, Nitsch H (2020) Cyber intelligence against radicalisation and violent extremism. In: Babak A, Douglas W, Blanco JM (eds) *Investigating radicalization trends: case studies in Europe and Asia*. Springer International Publishing, Cham, pp 55–80
8. Bogle A (2018) Strava has published details about secret military bases, and an Australian was the first to know. ABC News, 30 January. <http://www.abc.net.au/news/science/2018-01-29/strava-heat-map-shows-military-bases-and-supply-routes/9369490>
9. Burmaoglu S, Saritas O, Yalcin H (2019) Defense 4.0: Internet of things in military. In: *Emerging technologies for economic development*. Springer, Heidelberg, pp 303–320
10. Chapman E, Uren T (2018) *The Internet of insecure things*. Australian Strategic Policy Institute, Canberra
11. Chu G, Apthorpe N, Feamster N (2019) Security and privacy analyses of Internet of things children's toys. *IEEE Internet Things J* 6(1):978–985. <https://doi.org/10.1109/JIOT.2018.2866423>
12. Coady CAJ (Tony) (2004) Defining terrorism. In: Primoratz P (ed) *Terrorism: the philosophical issues*, pp 3–14. Palgrave, Basingstoke
13. Coady, CAJ (Tony) (2008) *Morality and political violence*. Cambridge University Press, Cambridge
14. Dang LM, Piran Md, Han D, Min K, Moon H (2019) A survey on Internet of things and cloud computing for healthcare. *Electronics* 8(7):768
15. de Haan W (2008) Violence as an essentially contested concept. In: Body-Gendrot S, Spiereburg P (eds) *Violence in Europe*. Springer, Heidelberg
16. Droogan J, Waldek L (2016) Where are all the cyber terrorists? From waiting for cyber attack to understanding audiences. 2016 Cybersecurity and cyberforensics conference (CCC), Aug. 2016, pp 2–4
17. Fernández-Caramés T, Paula F (2020) Teaching and learning IoT cybersecurity and vulnerability assessment with Shodan through practical use cases. *Sensors* 20(11). doi: <https://doi.org/10.3390/s20113048>
18. Fletcher D (2015) Internet of things. In: Blowers M (ed) *Evolution of cyber technologies and operations to 2035*. Springer, Dordrecht, pp 19–32

19. Frazer E, Hutchings K (2019) Can political violence ever be justified? Polity Press, Cambridge
20. Friedman B, David Hendry G (2019) Value sensitive design: shaping technology with moral imagination. MIT Press, Cambridge
21. Ghosh I (2020) AIoT: when artificial intelligence meets the Internet of things. *Visual Capitalist*, 12 August
22. Gill P, Horgan J, Hunter ST, Cushenbery LD (2013) Malevolent creativity in terrorist organizations. *J Creat Behav* 47(2):125–151. <https://doi.org/10.1002/jocb.28>
23. Habibzadeh, H, Dinesh K, Shishvan OR, Boggio-Dandry A, Sharma G, Soyata T (2019) A survey of healthcare Internet-of-things (HIoT): a clinical perspective. *IEEE Internet Things J* 7(1):53–71
24. Haynes J, Ramirez M, Hayajneh T, Bhuiyan MZA (2017) A framework for preventing the exploitation of IoT smart toys for reconnaissance and exfiltration. International conference on security, privacy and anonymity in computation, communication and storage
25. Henschke A (2017) Ethics in an age of surveillance: virtual identities and personal information. Cambridge University Press, New York
26. Henschke A (2017b) The Internet of things and dual layers of ethical concern. In: Patrick L, Keith A, Ryan J (eds) *Robot ethics 2.0: from autonomous cars to artificial intelligence*. Oxford University Press, Oxford
27. Henschke A (2020) Trust and resilient autonomous driving systems. *Ethics Inform Technol* 22:81–92. <https://doi.org/10.1007/s10676-019-09517-y>
28. Henschke A, Ford SB (2016) Cybersecurity, trustworthiness and resilient systems: guiding values for policy. *J Cyber Policy*, 1–14. doi: <https://doi.org/10.1080/23738871.2016.1243721>
29. Ingram HJ (2014) Three traits of the Islamic State's Information Warfare. *RUSIJ* 159(6):4–11. <https://doi.org/10.1080/03071847.2014.990810>
30. Ingram HJ (2015) The strategic logic of Islamic state information operations. *Aust J Int Aff* 69(6):729–752. <https://doi.org/10.1080/10357718.2015.1059799>
31. Ingram HJ (2017) An analysis of inspire and Dabiq: lessons from AQAP and Islamic State's Propaganda War. *Stud Confl Terror* 40(5):357–375. <https://doi.org/10.1080/1057610X.2016.1212551>
32. International Data Corporation (2015) Explosive Internet of things spending to reach \$1.7 trillion in 2020. According to IDC
33. Jenkins R (2013) Is Stuxnet real? Does it matter? *J Mil Ethics* 12(1):68–79
34. Ji W, Xu J, Qiao H, Zhou M, Liang B (2019) Visual IoT: enabling Internet of things visualization in smart cities. *IEEE Netw* 33(2):102–110
35. Kroes P (2009) Technical artifacts, engineering practice, and emergence. In: Krohs U, Kroes P (eds) *Functions in biological and artificial worlds: comparative philosophical perspectives*. MIT Press, Cambridge
36. Laqueur W (1977) *A history of terrorism*. Transaction Publishers, New Brunswick
37. Lazarescu M (2016) Hacked by your fridge: the Internet of things could spark a new wave of cyber attacks. *The Conversation*, 7 October
38. Manjikian M (2017) *Cybersecurity ethics: an introduction*. Routledge, London
39. Matyszczuk C (2015) Samsung's warning: our smart TVs record your living room chatter. CNet, February 8. Accessed 20 April 2016. <http://www.cnet.com/news/samsungs-warning-our-smart-tvs-record-your-living-room-chatter/>
40. Mayer M, Baeumner AJ (2019) A megatrend challenging analytical chemistry: biosensor and chemosensor concepts ready for the Internet of things. *Chem Rev* 119(13):7996–8027
41. Meneghello F, Calore M, Zucchetto D, Polese M, Zanella A (2019) IoT: Internet of threats? A survey of practical security vulnerabilities in real IoT devices. *IEEE Internet Things J* 6(5):8182–8201. <https://doi.org/10.1109/JIOT.2019.2935189>
42. Miller V, Hayward KJ (2018) 'I did my bit': terrorism, tarde and the vehicle ramming attack as an imitative event. *Br J Criminol* 59(1):1–23. <https://doi.org/10.1093/bjc/azy017>
43. Newman P (2020) The Internet of things 2020: here's what over 400 IoT decision-makers say about the future of enterprise connectivity and how IoT companies can use it to grow revenue. *Business Insider*, 7 March. <https://www.businessinsider.com/internet-of-things-report?IR=T>

44. Peresin A, Cervone A (2015) The Western Mujahirat of ISIS. *Stud Confl Terror* 38(7):495–509. <https://doi.org/10.1080/1057610X.2015.1025611>
45. Pramanik PKD, Pal S, Choudhury P (2018) Beyond automation: the cognitive IoT. Artificial intelligence brings sense to the Internet of things. In: Arun Kumar S, Thangavelu A, Meenakshi Sundaram V (eds) *Cognitive computing for big data systems over IoT: frameworks, tools and applications*, pp 1–37. Springer International Publishing, Cham
46. Primoratz I (2004) What is terrorism? In: Igor Primoratz (ed) *Terrorism: the philosophical issues*, pp 15–27. Palgrave, Basingstoke
47. Redden M (2016) Tech company accused of collecting details of how customers use sex toys. *The Guardian*, 14 September. <https://www.theguardian.com/us-news/2016/sep/14/wevibe-sex-toy-data-collection-chicago-lawsuit>
48. Rid T (2013) *Cyber war will not take place*. Hurst & Company, London
49. Robbins S (2019) A misdirected principle with a catch: explicability for AI. *Minds Mach* 29(4):495–514. <https://doi.org/10.1007/s11023-019-09509-3>
50. Robbins S, Henschke A (2017) Designing for democracy: bulk data and authoritarianism. *Surveill Soc* 15(3):582–589
51. Schiffer Z (2019) Smart TVs are data-collecting machines, *New Study Shows*. *The Verge*, 11 October. <https://www.theverge.com/2019/10/11/20908128/smart-tv-surveillance-data-collection-home-roku-amazon-fire-princeton-study>.
52. Schmitt MN (ed) (2013) *Tallinn manual on the international law applicable to cyber warfare*, Cambridge.
53. Song H, Rawat DB, Jeschke S, Brecher C (eds) (2017) *Cyber-physical systems: foundations, principles and applications*. Elsevier, London
54. Thebault R (2019) Woman's stalker used an app that allowed him to stop, start and track her car. *Washington Post*, 6 November. <https://www.washingtonpost.com/technology/2019/11/06/womans-stalker-used-an-app-that-allowed-him-stop-start-track-her-car/>.
55. West E (2019) Amazon: surveillance as a service. *Surveill Soc* 17(1/2):27–33
56. Zimmerman MJ (1997) Moral responsibility and ignorance. *Ethics* 107(3):410–426

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Facial Recognition for Counter-Terrorism: Neither a Ban Nor a Free-for-All



Scott Robbins

## 1 Introduction

This chapter starts from the fact that new technology has given new power to the state to automate the identification of previously known terrorists who are organizing attacks on the citizens that the state is supposed to protect. The power to do this (and associated powers), if it works effectively, would help in countering terrorism. Facial recognition technologies (FRTs) promise to give the state precisely that power.

Using FRTs, it is claimed, the state could *verify* that people are who they say they are, *identify* people appearing in images or video feeds, *characterize* their behavior and emotions, and check that they are not a suspected terrorist. For example, FRTs are deployed to verify that a person going through border control is indeed the person pictured on an identification document (e.g., a passport).<sup>1</sup> Interpol has deployed its Project FIRST system to help state authorities identify foreign terrorist fighters (FTFs).<sup>2</sup> In a truly horrifying example, the company Faception claims to be able to detect terrorists and pedophiles based on the characteristics of their face [16].<sup>3</sup> This power, however, has been challenged. This challenge, for some, should result in a complete ban on the use of this technology.

---

S. Robbins (✉)

Center for Advanced Security, Strategic and Innovation Studies (CASSIS), University of Bonn,  
Bonn, Germany

e-mail: [srobbins@uni-bonn.de](mailto:srobbins@uni-bonn.de)

<sup>1</sup> The company Veridas, for example, advertises that their FRT system can be used at border controls. See <https://veridas.com/government-institutions/>.

<sup>2</sup> <https://www.interpol.int/en/Crimes/Terrorism/Identifying-terrorist-suspects>.

<sup>3</sup> It would take another paper to discuss the many failings of even proposing this. While this is probably being used in earnest by both the state and private corporations— it should not be. The science behind applications like these are pseudo-science and have many of the same characteristics as the now disreputable practice of phrenology [2].

The reasons put forward for such a ban are that FRTs suffer from pervasive bias resulting in the benefits and harms being unequally distributed amongst groups, the state will inevitably use these technologies for illegitimate purposes, and that the existence of FRTs chill our behavior (i.e., causes people to censor themselves for fear of surveillance).

At the moment, the state faces little restriction over how they use FRTs. There are plenty of examples of the state's use of FRTs for purposes that give people pause. For example, police departments in the U.S. have used FRTs to identify and monitor activists and protestors of color [1]. Setting up a surveillance network powered by FRTs will, it is argued, significantly increase the risk of the state abusing its power. This risk will be associated with an increase in citizens chilling their behavior.

It is paramount that if the state can use this technology to increase their power to counter terrorism, this power is constrained such that the abuses and chilled behavior do not occur. I argue below for five conditions on the use of FRTs. First, the state must create institutional constraints that only allow FRTs to be used in places where people do not (and should not) enjoy a reasonable expectation of privacy (e.g., airports, border crossings). Second, the cameras equipped with FRT must be marked to assure the public that they are not being surveilled in places that they should have a reasonable expectation of privacy. Third, FRTs should be restricted to finding serious criminals (e.g., terrorists). Fourth, the state should not use third-party companies that violate the first three conditions during the creation or use of its service. And fifth, third-party companies should not be able to access or read the sensitive data collected by the state. With these conditions satisfied, given the effectiveness of FRT, the state can harness FRT's power to counter-terrorism.

## 2 The Basics of Facial Recognition

The goal of FRT is to verify, identify, characterize, or check someone against a watch list based on an image of a particular person. Most people will have some experience with this because Facebook, Google, and Apple all use F.R. in commercial applications. Apple's FaceID lets users into their phones using the camera on their phone to verify their identity. Google and Facebook have long used F.R. to identify and auto-tag photos with the people in them.

The four goals of FRTs (verify, identify, characterize, and watch list)<sup>4</sup> should be distinguished as they carry different ethical concerns. Verification is merely matching two images to check if they are the same person. This is a one-to-one comparison. An organization may want to verify that the person wearing a security badge is the same as the person's picture on the badge. This can be difficult for humans to do—but relatively easy for FRTs.

---

<sup>4</sup> There is also a fifth goal that is simply *detection* which would merely involve detecting that there is a face in an image. This is necessary in order for the success of the other goals but needn't concern us here.

Using FRTs to identify is a one-to-many relationship. An image of a face has to be checked against a ‘faces’ database to determine who, exactly, they are.<sup>5</sup> This goal of FRTs has had its spotlight in the media recently because the FBI used FRT to identify those who participated in the January 6, 2021 insurrection at the U.S. capitol [6]. High-quality images taken from that day are fed into an FRT that can compare the faces on those images to a database of faces that have identities attached to them.

FRTs for the characterization of a particular person aim to label people as having a particular emotional state or as, for example, terrorists based solely on an image of their face. In China, for example, FRTs have been used in classrooms to detect the level of engagement of the students [12]. The aforementioned company Faception claims to be able to detect everything from ‘professional poker player’ to a person with a ‘high IQ’ to a ‘terrorist’. This has been dismissed as modern day phrenology by some academics—as there does not seem to be any evidence that facial features have a relationship to personality, profession, or criminal behavior (see e.g., [28]).

### 3 Arguments for an FRT Ban

#### 3.1 *Disparate Impact*

FRTs suffer from pervasive bias. This means that FRTs perform exceptionally well for some groups, while it performs terribly for other groups. With this in mind, we can use ‘bias’ how we use it in everyday language: FRTs are biased against dark-skinned people. FRTs, in one study, performed 5–10% worse for African Americans compared with Caucasians. Buolamwini and Gebru found error rates as high as 34% for African females compared with a low 0.8% error rate for Caucasian males. This could be due to their being a lack of images of dark-skinned people used to train the algorithm. Or it could be that dark-skinned faces are harder for current algorithms to translate into computer language and extract useful patterns out of. Whatever the reason, current algorithms have a huge problem recognizing dark-skinned faces.

Problems like these mean that the benefits and harms brought by FRTs are unequally distributed amongst groups of people. Those for whom the technology does not work as well with, will not be able to be verified by FRTs—causing suspicion and further intrusive surveillance. Furthermore, they will be misidentified more frequently. This may cause them to be suspected as a terrorist or other serious criminal. For example, on January 9, 2020, Robert Williams was arrested in front of his wife and two daughters. The reason for this arrest was that an FRT misidentified him as a person who stole watches from a store in a robbery that took place 18 months earlier [15]. If Robert Williams was just an unfortunate misidentification due to the FRT not being 100% accurate, we could accept this—misidentification also happens

---

<sup>5</sup> I am speaking loosely here. The ‘faces’ in the database are actually computer generated representations of faces. Depending upon the specific methodology used these can be more or less robust. For an overview of some of the specific technical methodologies see [11].

when done by human beings. However, FRTs consistently misidentify (or fail to identify) people of color significantly more often than whites. Then, it follows that people of color will disproportionately experience the harms caused by these technologies. In a liberal democracy, the principle of everyone being equal under the law is violated by technologies with this problem.

Meanwhile, the benefits of FRTs will be disproportionately received by middle-aged white males. Not only will they be identified more reliably—meaning that they will get through security lines without further intrusive surveillance, but they will disproportionately feel the benefits of convenience that these technologies promised in the first place. Joy Buolamwini (mentioned above) started to analyze FRTs precisely because she couldn't get FRTs to recognize her face. At one point, she put on a white mask triggering the program to recognize hers as a face [13]. The point is that the convenience promised by FRTs is distributed unfairly. This compounds the problem above because the same group that disproportionately experiences the harms of FRTs also disproportionately fail to experience its benefits. This problem must be overcome if FRTs are to be used anywhere. The main point is that many FRTs don't work. If a particular technology doesn't work, then we shouldn't use it. However, this does not mean that the technology will not work in the future. In this paper, I assume that we will only be using FRTs that work with an appropriate level of effectiveness for everyone.

### 3.2 *Chills Behavior*

*Surveillance conducted with facial recognition systems is intrinsically oppressive. The mere existence of facial recognition systems, which are often invisible, harms civil liberties, because people will act differently if they suspect they're being surveilled [7].*

Many institutions and scholars echo this sentiment about FRT. Evan Selinger and Brenda Leong, channeling philosopher Benjamin Hale, argue that pervasive effective FRTs would undermine our free will—and would prevent ethical behavior caused by that will—replacing it with “I acted ethically because someone was watching” [22]. The freedom to choose to do the right thing whether or not someone is watching is central to the liberal democratic ideal of autonomy. In everyday life we encounter many scenarios that require ethical reasoning and action. Coffee, for example, might be for sale based on the honor system. Customers are supposed to leave a Euro after they take a coffee. People should have the right to be honorable. When someone (or something) is watching, then we don't get the chance to be honorable. Our actions are evaluated in light of someone watching—which, when you leave the Euro for the coffee isn't as honorable as if you were to leave the Euro without someone watching. FRTs, therefore, should be banned (or so concludes their argument).

Furthermore, the freedom to gather in large groups to protest injustice should not be hampered by the knowledge that you will be identified by FRTs and be labeled as a subversive. The freedom of assembly is enshrined in liberal democratic



constitutions and declarations of human rights. The U.N. Declaration of Human Rights, in article 20, states that “Everyone has the right to freedom of peaceful assembly and association” [25], and the United States Constitution gives citizens the “right of the people peaceably to assemble, and to petition the Government for a redress of grievances (“U.S. Senate: Constitution of the United States”, n.d. 26).” The right to assemble and air grievances can be the last resort to create necessary change. A 2020 study of the U.S. civil rights movement, for example, showed that it was activism and protests which “drove media coverage, framing, congressional speech, and public opinion on civil rights” [27].

A person who was horrified by the murder of George Floyd and wants to voice their support for systemic change in the policing system in the U.S. should be able to do so. However, they may fear that FRTs will identify them as taking part in a protest (and may further document what exactly the protest was)—which may cause them to lose their job or harm their chances for jobs in the future. If that protest were to turn violent then it may be that all attendees get labeled as violent protestors—regardless of their intentions and actions at the protest. A 2013 report showed that:

*surveillance of Muslims’* quotidian activities has created a pervasive climate of fear and suspicion, encroaching upon every aspect of individual and community life. Surveillance has chilled constitutionally protected rights—curtailing religious practice, censoring speech and stunting political organizing [23].

Ordinary citizens’ rights to practice religion, speak their minds, and politically organize have been shown to be hindered by surveillance. FRTs increase this risk dramatically—as their chances of being identified with FRTs is far greater.

The recent events of January 6, 2021, in which pro-Trump groups converged on the capital and staged a violent insurrection, may cause one pause here. Don’t we want these people to have their behavior ‘chilled’? Many liberals cheered the use of FRTs to identify people to have them arrested.

There are two essential things to note here. First, the right to assembly, speech, and political organizing does not include the right to violently overthrow the government. Chanting and holding up signs in front of the Capitol building should not be chilled—whether or not we agree with the assemblers. Carrying weapons, engaging with police, and threatening Congress members are not included in the right to peaceful assembly. Second, while this protest did turn violent and illegal, that does not mean that each individual who attended this protest deserves to be stigmatized without participating in the actual insurrection. While those that stormed the capitol should fear consequences brought on by the state, those that simply protested the election results should not.

If people who merely intended to voice their grievances did not attend this protest simply because they feared being identified and face the consequences, their right to peaceful assembly was violated. This causes harm even if that person is wrong about what might happen to them. They are unsure—and therefore change their behavior. This is why there must be both institutional barriers to technology being used this

way *and* transparency in law enforcement and government to assure the public that this is so [20].<sup>6</sup>

One might think that with CCTV we already face this problem. CCTV captures images of people all the time—and sometimes that footage is distributed in order to identify someone that has committed a crime. Think of an armed robbery at a gas station. The suspect might be captured on CCTV footage in front of the gas station—and could be captured because someone recognized them on that footage. If, two days later, someone else walks into that same gas station they are also captured on CCTV footage. However, because normal CCTVs are not equipped with FRTs, they will never be identified as there was no crime committed (the footage has no reason to be ‘looked’ at—by a computer or a human). FRTs have the capability to continuously identify and store the information related to people that come across its view. This affords the state the ability to easily identify anyone who attended a particular protest—whether or not they committed a crime. This amounts to intrusive surveillance without cause. Of course, the state can claim that they do not store the information unless crimes are committed; however without clear and transparent institutional (and possibly technological) barriers to such use, it will be difficult for people to act as if they are not being surveilled using FRTs.

### 3.3 *Scope Creep*

*Facial recognition enables surveillance that is oppressive in its own right; it's also the key to perpetuating other harms, civil rights violations, and dubious practices. These include rampant, nontransparent, targeted drone strikes; overreaching social credit systems that exercise power through blacklisting; and relentless enforcement of even the most trivial of laws, like jaywalking and failing to properly sort your garbage cans [21].*

The arguments for a ban rest on the premise that FRTs will creep pervasively into society and be used for all kinds of things they weren't initially used for. In the above quote, FRTs are envisioned for Chinese-style social credit systems and enforcement of things like jaywalking. The idea is that once this technology is out there, it will be normalized. We will come to expect it—and then it will be used everywhere.

To highlight this, I offer the following example. Let's say that FRTs are extremely effective. The government has intelligence that five New York City individuals are planning on carrying out a terrorist attack. It is decided to upgrade the CCTV network to include FRT. If any one of those individuals is captured by the smart CCTV cameras, then the authorities will be notified. It is agreed that this will be their best chance to stop the terrorist attack. Unfortunately, that upgrade cost a lot of money. In an attempt to raise money, the mayor decides that the FRT can simply start automatically ticketing J walkers. J walking is illegal, and many people do it—so using FRT to ticket them will raise a lot of money.

---

<sup>6</sup> More on this in Sect. 4.1.

Jeroen van den Hoven calls this ‘information injustice.’ He argues that people may not object to their data being used for a particular purpose; however, when that same data is used for another purpose, an injustice has occurred. If your library search data is collected to provide better services by the library, this may be something you agree to. However, if that same data is used to collect information on your tastes and pass them to others for advertising purposes, then informational injustice has occurred [10]. In this case, the use of FRTs to catch terrorists is now repurposed for catching J walkers. While an argument may justify the use of FRTs to catch terrorists, it cannot be used to catch J walkers without a new justification.

The problem is that once the surveillance apparatus includes pervasive FRTs, the barriers to using it for things not originally intended are very low. This is not the case for regular CCTV cameras. The cost of employing people to pour through that video and attempt to identify individual J walkers wouldn’t be worth the money raised by ticketing them. CCTV’s technological limitations naturally restrict law enforcement’s ability to use them for anything—protecting people’s reasonable expectation of privacy.

### ***3.4 An Outright Ban***

For some, the concerns above, taken together, creates a case for an outright ban of the technology. Like San Francisco, some cities have enacted such a ban [3].

Selinger and Werner believe that FRTs are “so inherently toxic that it deserves to be completely rejected, banned, and stigmatized” [21]. In another post, they conclude that “The future of human flourishing depends upon facial recognition technology being banned before the systems become too entrenched in our lives” [7].

In what follows, I argue that an outright ban may not be justified. First, there are contexts in which our expectation of privacy is simply non-existent. Second, the chilling effects are not necessarily going to happen, nor are they necessarily bad things. Finally, the scope creep that critics are concerned about is not inevitable. If the technology works as advertised, then there are some restricted contexts where these harms do not materialize.

## **4 Conditions for the Use of Facial Recognition**

Given the argument for bans on FRT and the privacy and free speech rights enshrined in liberal democratic constitutions and human rights declarations, it is clear that the state must justify the use of FRTs before they can be used to capture terrorists. This is not a technology that simply improves upon a power that the state already had; instead, it is an entirely novel power. That is the power to identify anyone that comes into view of an FRT equipped camera without a human being watching the video feed.

Here I will outline the conditions that FRT should be subject to operate in a liberal democracy justifiably. I expand on each in the sections below. The context in which FRT is being used must be one in which the public does not have a reasonable expectation of privacy. Second, the only goal should be to prevent serious crimes like terrorism from taking place. Finally, FRTs to store and capture biometric facial data in a database, the individual in question must be suspected of committing a serious crime.

#### ***4.1 Reasonable Expectation of Privacy***

In a famous case in the United States, the supreme court ruled that Charles Katz had a reasonable expectation of privacy when he closed the phone booth door [4, Chap. 1]. This meant that the evidence collected by the state who was listening in on his conversations in that phone booth had to be thrown out. This notion of a ‘reasonable expectation of privacy’ is fundamental to how the value of privacy is interpreted in liberal democracies. It is not just a legal notion but a notion which grounds how we act. In our bedrooms, we have a reasonable expectation of privacy, so we can change clothes without fear of someone watching. When Charles Katz closed the door to the phone booth he was using, he enjoys a reasonable expectation of privacy—he believes that no one should listen to his conversation.

Facial data captured by FRTs should be at least as protected as voice data. CCTVs in the public sphere should not be collecting information on individuals—something that happens when CCTVs are equipped with FRT. When I walk down my street, I have a reasonable expectation that my comings and goings are not being recorded—whether it be a police officer following me around or by a smart CCTV camera recognizing my face. Regular CCTVs do not record individuals’ comings and goings; rather, they record what happens at a particular location.

The difference is that a CCTV camera does not record a line in a database that includes my identity and the location that I was ‘seen’ at. CCTV equipped with FRT *can* record such a line in a database—significantly empowering the state to perform searches that tell them much about my comings and goings. Not only should these searches be linked to clear justifications; but there should be clear justifications for collecting such intimate data (their comings and goings) on individuals.

This reasonable expectation can be overridden if I have committed a serious crime or plan on committing a serious crime. This is because my right to privacy would be overridden by the “rights of other individuals...to be protected by the law enforcement agencies from rights violations, including murder, rape, and terrorist attack” [17, 110]. If one were to be in the process of planning a terrorist attack, it would not be a surprise to them that they were being surveilled. Terrorists take active measures to prevent surveillance that they expect to occur. This may seem to justify the placing of smart CCTVs in public spaces to identify terrorists.

CCTV cameras are currently placed in many public spaces. If something happens, the authorities can review the CCTV footage to see who was responsible. In this case,

the place itself is being surveilled. Data on individuals is not ‘captured’ in any sense. There is no way to search a database of CCTV footage for a particular name. One must look at the footage. However, if this CCTV camera were to be “smart” and capture biometric facial data along with video footage, then each individual who is captured by this camera is being surveilled. The authorities now know each person that comes into this camera’s view and what time they were there. This, even though an overwhelming majority of people coming into any CCTV camera’s view has not, and does not plan to, commit a serious crime. Their privacy has been invaded.

This has ethical implications regarding scope creep and chilling behavior discussed in Sect. 3. If FRT enabled CCTV cameras are in operation, then it is easy for the state to add new uses for the technology. A simple database search could reveal everyone who goes into an area with many gay bars. A gay man in a country where homosexuality is considered unacceptable but not illegal may chill their behavior—that is, not go to gay bars to fear those visits being documented. While the FRT enabled CCTV cameras were initially installed to counter terrorism, the ability to easily search for anyone that has come across it makes it easy to use it for other, illegitimate purposes.

The state could simply state that they will only use FRTs with a warrant targeted against an individual suspect of a serious crime. For example, the authorities may have good information regarding the planning of a terrorist attack by a particular person. It is imperative that they find this person before they are able to execute the attack. They obtain a warrant and then use the city’s network of FRT-enabled CCTV cameras to ‘look’ for this person. If this person’s face is captured by one of these cameras, then the authorities are immediately notified.

If we bracket issues of efficacy and disparate impact, it appears that this would be a useful power to the state—and subject to restrictions that protect privacy. The issue is not whether or not to use FRTs, but *how* they can and should be used. However, these would be merely institutional and perhaps legal barriers that are subject to interpretation. The scope of national security is little understood. Donald Trump used the concept to justify the use of collecting cell-phone location data to track suspected illegal immigrants [14]. The power enabled by FRTs is so great, and the justifications to use them will be so little understood, that it will be near impossible for regular citizens to feel and act as if they have privacy—even if they do, in principle, have it. Your partner may promise to never read your journal unless you are either dead or in a coma; however, the fact that she has a key and knows where it is will probably cause you do self sensor what you write down—just in case. With a journal, and with your general comings and goings, you should enjoy a reasonable expectation of privacy.

However, there are some public spaces where individuals do not enjoy a reasonable expectation of privacy. Airports and border crossings are two such examples. For better or worse, we now expect little privacy in these contexts. Authorities are permitted to question us, search our bags, search our bodies, submit us to millimeter scans, etc. It would be rather odd to think that our privacy was invaded more by our faces being scanned and checked against a criminal database. On regular public sidewalks, I would be horrified to find out that the state recorded my comings and

goings; however, I would be shocked to find out the state did not record each time I crossed into and out of the country. This points to the idea that there may be places where we *should* have a reasonable expectation of privacy—whether we do or not.

A recent U.S. supreme court case illustrates this nicely. Timothy Carpenter was arrested for armed robbery of Radio Shacks and T-Mobile stores. The police used a court order (which is subject to less standards than a warrant) to obtain GPS data gathered by his cell phone and collected by the telecommunications companies MetroPCS and Sprint. In an opinion written by chief justice John Roberts, the supreme court ruled that Timothy Carpenter should have a reasonable expectation of privacy concerning his constant whereabouts. The government cannot simply, out of curiosity, obtain this data [24]. This prevents the widespread use of smart CCTV cameras in plain sight to undermine our ‘reasonable expectation of privacy.’ The state should not use conspicuous surveillance as a way to claim that no one has a reasonable expectation of privacy where these cameras exist. The critical point is that there are public spaces where citizens of a liberal democracy *should* have a reasonable expectation of privacy.

Therefore, *if* there are places where citizens *should not* have a reasonable expectation of privacy *and* FRTs are effective (they do not cause unequally distribute false positives and false negatives across different groups), it may be justifiable to use FRTs in those places. People expect the state to protect them from terrorism. If FRTs contribute to keeping citizens safe from terrorists, then there is a good reason to use them. However, based on the analysis above, they cannot simply be used anywhere as there are places where citizens *should* have a reasonable expectation of privacy.

The above points to the allowable use of regular CCTV cameras in public spaces but prevents FRTs from operating in those same public spaces.<sup>7</sup> The problem now is: How will the public know the difference? This is a serious problem. After all, the right to free expression may be ‘chilled’ because people believe that the state is surveilling their actions. I may worry that because my friend lives above a sex shop, the state’s surveillance may cause them to believe I frequent the sex shop rather than visit my friend. I may, therefore, not visit my friend very often. Or I may not join a Black Lives Matter protest because I believe the state is using FRTs to record that I was there. This is the “chilling effect” mentioned in Sect. 3.2. This can occur even if the state is *not* engaging in such surveillance. The only thing that matters is that I believe it to be occurring.

The ‘chilling effect’ puts the burden on the state to assure the public that such unjustified surveillance is not happening. Where it is justified, there are appropriate safeguards and oversight to prevent misuse, etc. This requires institutional constraints, laws, and effective messaging. As [20] argue, institutional constraints and laws alone will not *assure* the public that the state is not practicing unjustified

---

<sup>7</sup> It is not for this paper to evaluate the ubiquitous use of regular CCTV cameras in public spaces. I only claim that regular CCTV does not violate our reasonable expectation of privacy if it is the place that is being surveilled and not individual people (e.g. when our identities, time, and location are stored in a searchable database).

intrusive surveillance. And vice versa, effective messaging alone will not *ensure* that the state is not practicing unjustified intrusive surveillance.

For example, if the state creates laws that prevent the use of FRT on regular city streets but the cameras that are used look the same as the smart CCTV cameras that have FRT in airports, then the public will not be assured that facial recognition is not taking place. This sets up the conditions for the chilling effect to occur. However, if the state uses cameras that are clearly marked for facial recognition in places like airports, and cameras that are clearly marked ‘no facial recognition’ on city streets but no laws are preventing them from using FRT on city streets, then the public has a greater chance of being assured. However, nothing is preventing the state from using the footage of those cameras and running facial recognition on them after the video has been captured. Therefore, it takes both institutional constraints (bound by law) *and* effective messaging to meet the standards which support liberal democratic values like free expression.

This creates two conditions for the state’s use of FRT. First, the state must create institutional constraints that only allow FRTs to be used in places where people do not (and should not) enjoy a reasonable expectation of privacy (e.g., airports, border crossings). Second, the cameras equipped with FRT must be marked to assure the public that they are not being surveilled in places that they should have a reasonable expectation of privacy.

#### ***4.2 Cause for the State’s Use of FRTs***

The state should not simply use new technology because it exists. There must be a purpose for using technology that is greater than the harms and privacy infringements that occur due to that technology. It would be odd to use wiretaps to surveil a serial jaywalker. Wiretaps are used in highly restrictive situations involving serious criminals. FRTs should be no different. The point is, that “justifications matter.” Collecting facial data by using FRTs for countering terrorism does not mean that the data is now fair game for any other use. Each use must have its moral justification—and if that justification no longer obtains, then that data should be destroyed [8, 257].

Terrorism is a serious enough risk (in terms of possible harm—not necessarily in terms of likelihood) that it features as a justification employed by those advocating the use of FRTs. In these cases, one does not feel as if the privacy rights of terrorists are so strong that they should not be surveilled. We expect the government to do what they can to find people like this. Their privacy rights are overridden by others’ rights not to be injured or killed in a terrorist attack.

The problem is that FRTs must also surveil everyone that comes into view of one of its cameras. That is, each face is used as an input to an algorithm that attempts to match that face to an identity and/or simply check whether that face matches one of the identities of suspected terrorists. In a technical sense, this technology could only be used for the legitimate purpose of finding terrorists. However, as argued above—the difficulty in assuring the public that this is the case will have a chilling

effect. Furthermore, the real possibility of scope creep makes placing these cameras, in places where people should have a reasonable expectation of privacy, dangerous.

This means that no matter the cause, FRTs should not be employed in places where innocent people have a reasonable expectation of privacy (as argued above). However, once we restrict its use to those places where there is no reasonable expectation of privacy, then finding serious criminals using FRTs poses no ethical problem (providing that it reaches a threshold of effectiveness). The third condition for the use of FRTs is that FRTs should be restricted to finding serious criminals (e.g., terrorists).

### ***4.3 Reliance on Third-Party Technology***

The state's reliance on third-party technology companies to facilitate surveillance is perhaps the area where the most violations of liberal democratic values occur. For example, the government cannot simply scrape the entire internet of pictures of people, match the faces to names, create a detailed record of things you have done, places you have gone, people you have spent time with, etc. Especially without a just cause. This amounts to intrusive surveillance of every individual. In liberal democracies, there must be a justification (resulting in a warrant approved by a judge) to engage in such surveillance of an individual. Surveilling a million people should not be considered more acceptable than the surveillance of one person. However, Clearview A.I. has been scraping images from the web and creating digital identities for years. Many police departments and government agencies are now using this third-party company to aid in using FRTs [9].

This causes significant ethical concern for three reasons: first, some third-party companies do not follow the constraints already mentioned above; second, sensitive data is being stored and processed by third-party companies that have institutional aims that could incentivize the misuse or abuse of this data; and third, the role that these companies play in surveillance may reduce the public's trust in them.

#### **4.3.1 Contracting out the Bad Stuff**

When I first encountered FRT at an airport, I was a bit squeamish. It took me some time to understand why. Indeed, I am not against using such technology to prevent terrorists from entering the country or detecting people who are wanted in connection with a serious crime or find children on missing person lists.<sup>8</sup> I also did not feel that I had a reasonable expectation of privacy. I expect to be questioned by a border guard and have my passport checked. I expect that my bag or my body could be searched.

---

<sup>8</sup> Although I was concerned that my face could be checked against those captured at, for example, Black Lives Matters protests around the world—and that I could face scrutiny due to my participation. This concerns the just causes for FRTs discussed earlier.



And I expect to be captured on camera continuously throughout the airport. So why did I have this immediate adverse reaction towards the use of FRT by the state?

The answer lies in my knowledge regarding the contracting out of such work to third-party technology companies. I am expected to trust the state and the third party technology company that is behind the technology. Are they capturing my biometric face data and storing it on their third-party servers? Are there institutional barriers preventing them from reusing or selling that data for their benefit? Is the data captured, sent, stored, and processed in line with best security practices? In short, I fear that even if the proper laws and constraints regarding the state's use of FRTs are in place, that third-party technology company is not bound by them or does not respect them.<sup>9</sup>

This is wrong. There are laws in place that prevent the United States, for example, contracting out intrusive surveillance on their citizens to other countries. So the U.S.—not being able to collect data on its citizens—cannot ask the U.K. to collect data on a U.S. citizen. The same should be true for FRTs. Suppose the U.S. cannot gather facial data on the entire U.S. population (practicing bulk surveillance). In that case, the U.S. should also not contract such work out to a third-party company—or use a third party company that has engaged in this practice. If I contract the work of killing an enemy to somebody else, that does not absolve me of all responsibility regarding the murder of that enemy.

It is not, in principle, unacceptable to use tools created by third-party companies. Third-party companies often have the resources and incentives to create far better tools than the government could create. Silicon valley technology companies attract many creative and motivated thinkers—and pay them a salary that the government could not afford. It would be detrimental to say that the government cannot use tools created by these companies. However, big data and artificial intelligence have made this relationship much more complicated.

Rather than merely purchasing equipment, the government is now purchasing services and data. A.I. algorithms created by third-party companies are driven by the collection of vast amounts of data. If this algorithm is to be used by the state, the state must ensure that the data driving it was collected according to laws governing the state's data collecting capabilities. Furthermore, the hosting of the data that the government collects is increasingly being contracted out to cloud services like Amazon Web Services. This is so because this data processing is extremely resource-intensive and something that third-party companies are more efficient at. This creates a situation where our biometric facial data may have to be sent to a third-party company for storage and/or processing. The company in question must have no ability to see/use this data. This is so for two reasons. First, these companies have institutional aims<sup>10</sup> that have nothing to do with the security of the state. This creates incentives for companies to use this data for their aims—creating an informational injustice [10]. Furthermore, this blurring of institutional aims (e.g.,

---

<sup>9</sup> It should be noted that strong data protection laws like Europe's GDPR can prevent some of this from taking place.

<sup>10</sup> See Miller [18] for an excellent discussion on the blurring of institutional purposes.

maximizing profits *and* countering terrorism) could be detrimental to the company. As a result of NSA programs like PRISM, which purportedly allows the state to gain access to the company servers of Google and Facebook [5], rival companies are now advertising that they are outside of U.S. jurisdiction and can therefore be used without fear of surveillance.<sup>11</sup>

Second, this data is now being entrusted to companies that may not have the same security standards or oversight expected for the storage and processing of sensitive surveillance data. Recently the Customs and Border Patrol contracted out facial recognition to a third-party company which was breached in a cyber-attack causing the photos of nearly 100,000 people to be stolen. Customs and Border Patrol claimed no responsibility—saying it was the third-party company’s fault. The state should be responsible for the security of surveillance data [19, 35].

This discussion should cause constraints on how the state uses third-party companies to facilitate surveillance. Condition number four for the state’s use of FRTs is that the state should not use third-party companies that violate the first three conditions during the creation or use of its service. This means that the state should know about the services they are using. Furthermore, a fifth condition is that the third-party company should not be able to access or read the sensitive data collected by the state. This keeps the state in control of this sensitive surveillance data.

## 5 Conclusion

What has been written above agrees with much of what proponents of a ban argue. The large difference is that I do not believe that FRTs will necessarily creep into society in a pervasive way. The five conditions I argue for above prevents this type of creep. Furthermore, the chilling effect so feared by proponents of a ban will not necessarily occur. This only happens when there is pervasive use of FRTs in places where people should have a reasonable expectation of privacy. By restricting FRTs use to those places where people should not have a reasonable expectation of privacy, this concern can be alleviated.

However, the concern that FRTs suffer from pervasive bias is serious. There may not be FRTs that are effective at all. This should prevent their use by the state. Until it can be shown that these technologies work in a way that won’t disproportionately distribute the harms and benefits amongst groups, FRTs should not be used. What is called for, then, is a moratorium rather than a ban. Once it has been shown that FRTs are effective, the state should use them within the limits outlined above.

---

<sup>11</sup> ProtonMail, for example, claims that “ProtonMail is incorporated in Switzerland and all our servers are located in Switzerland. This means all user data is protected by strict Swiss privacy laws.” <https://protonmail.com/>.

## References

1. Cagle M (2016) Facebook, Instagram, and Twitter provided data access for a surveillance product marketed to target activists of color. ACLU of Northern CA. October 11, 2016. <https://www.aclunc.org/blog/facebook-instagram-and-twitter-provided-data-access-surveillance-product-marketed-target>
2. Chinoy S (2019) Opinion. The racist history behind facial recognition. The New York Times, July 10, 2019, sec. Opinion. <https://www.nytimes.com/2019/07/10/opinion/facial-recognition-race.html>
3. Conger K, Fausset F, Kovaleski SF (2019) San Francisco bans facial recognition technology. The New York Times, May 14, 2019, sec. U.S. <https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html>
4. Farivar C (2018) Habeas data: privacy vs. the rise of surveillance tech. Brooklyn
5. Greenwald G (2015) No place to hide: Edward Snowden, the NSA, and the U.S. surveillance state, reprint. Picador, New York
6. Harris M (2021) How facial recognition technology is helping identify the U.S. capitol attackers - IEEE Spectrum. IEEE Spectrum: Technology, Engineering, and Science News. January 11, 2021. <https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/facial-recognition-and-the-us-capitol-insurrection>
7. Hartzog W, Selinger E (2018) Facial recognition is the perfect tool for oppression. Medium. August 2, 2018. <https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66>
8. Henschke A (2017) Ethics in an age of surveillance: personal information and virtual identities. Cambridge University Press, New York
9. Hill K (2020) The secretive company that might end privacy as we know it. The New York Times, January 18, 2020, sec. Technology. <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>
10. van den Hoven J (1999) Privacy of information injustice? In: Mendina GT (ed) Ethics and electronic information in the twenty-first century. Purdue University Press, West Lafayette, Indiana, USA, pp 139–150
11. Introna LD, Nissenbaum H (2009) Facial recognition technology: a survey of policy and implementation issues. The center for catastrophe preparedness and response. [https://nissenbaum.tech.cornell.edu/papers/facial\\_recognition\\_report.pdf](https://nissenbaum.tech.cornell.edu/papers/facial_recognition_report.pdf)
12. Kuo L (2019) China brings in mandatory facial recognition for mobile phone users. The Guardian, December 2, 2019, sec. World news. <http://www.theguardian.com/world/2019/dec/02/china-brings-in-mandatory-facial-recognition-for-mobile-phone-users>
13. Lohr S (2018) Facial recognition is accurate, if you're a white guy. The New York Times, February 9, 2018, sec. Technology. <https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html>
14. Lutz E (2020) Trump's immigration crackdown has taken a dystopian turn. Vanity Fair. February 7, 2020. <https://www.vanityfair.com/news/2020/02/trump-immigration-crackdown-has-taken-a-dystopian-turn-cell-phone-data>
15. Mayor P (2020) ACLU letter to detroit public safety headquarters, June 24, 2020. [https://cdn.arstechnica.net/wp-content/uploads/2020/06/dpd\\_complaint\\_v\\_final.pdf](https://cdn.arstechnica.net/wp-content/uploads/2020/06/dpd_complaint_v_final.pdf)
16. McFarland M (2016) Terrorist or pedophile? This start-up says it can out secrets by analyzing faces. Washington Post, May 24, 2016. <https://www.washingtonpost.com/news/innovations/wp/2016/05/24/terrorist-or-pedophile-this-start-up-says-it-can-out-secrets-by-analyzing-faces/>
17. Miller S (2008) Terrorism and counter-terrorism: ethics and liberal democracy. Wiley. <https://www.wiley.com/en-us/Terrorism+and+Counter+Terrorism%3A+Ethics+and+Liberal+Democracy-p-9781405139434>
18. Miller S (2019) Whither the University? Universities of technology and the problem of institutional purpose. Sci Eng Ethics 25(6):1679–1698. <https://doi.org/10.1007/s11948-019-00147-7>

19. Robbins S (2021) Machine learning & counter-terrorism: ethics, efficacy, and meaningful human control. Doctoral thesis, Delft. Technical University of Delft, The Netherlands. <https://repository.tudelft.nl/islandora/object/uuid:ad561ffb-3b28-47b3-b645-448771eddaff>
20. Robbins S, Henschke A (2017) The value of transparency: bulk data and authoritarianism. *Surveill Soc* 15 (3/4): 582–589. <https://doi.org/10.24908/ss.v15i3/4.6606>
21. Selinger E, Hartzog W (2018) Amazon needs to stop providing facial recognition Tech for the Government. Medium. June 21, 2018. <https://medium.com/s/story/amazon-needs-to-stop-providing-facial-recognition-tech-for-the-government-795741a016a6>
22. Selinger E, Leong B (2021) The ethics of facial recognition technology. SSRN scholarly paper ID 3762185. Rochester, Social Science Research Network, NY. <https://papers.ssrn.com/abstract=3762185>
23. Shamas D, Arastu N (2013) Mapping muslims: NYPD spying and its impacts on American Muslims. Long Island, New York, CUNY School of Law, USA. <https://www.law.cuny.edu/wp-content/uploads/page-assets/academics/clinics/immigration/clear/Mapping-Muslims.pdf>
24. Sorkin AD (2018) In carpenter, the supreme court rules, narrowly, for privacy. *The New Yorker*. June 22, 2018. <https://www.newyorker.com/news/daily-comment/in-carpenter-the-supreme-court-rules-narrowly-for-privacy>
25. Universal Declaration of Human Rights (2015) October 6, 2015. <https://www.un.org/en/universal-declaration-human-rights/>
26. U.S. Senate: Constitution of the United States. n.d. Accessed February 4, 2021. [https://www.senate.gov/civics/constitution\\_item/constitution.htm](https://www.senate.gov/civics/constitution_item/constitution.htm)
27. Wasow O (2020) Agenda seeding: How 1960s Black protests moved Elites, public opinion and voting. *Am Polit Sci Rev* 114(3):638–659. <https://doi.org/10.1017/S000305542000009X>
28. Whittaker M, Crawford K, Dobbe R, Fried G, Kaziunas E, Mathur V, West SM, Richardson R, Shultz J, Schwartz O (2018) AI Now 2018. AI Now Institute. December 2018. [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.html](https://ainowinstitute.org/AI_Now_2018_Report.html)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# The Rise of the Modern Intelligence State



John Hardy

**Abstract** The rise of the formal surveillance state in the early twenty-first century was precipitated by political impetus to empower security and intelligence organisations to perform a broad range of counterterrorism functions. Ethical debates about the implications of the security intelligence reach of modern states have focused on balancing individual rights, liberties, and privacy against the security of the state. Meanwhile, the surveillance state has rapidly evolved into an intelligence state, capable not only of pervasive data collection, but also of analytical modelling which expands existing boundaries of surveillance. Existing concerns about the ethical collection and use of surveillance data are compounded by three emergent capabilities of the modern intelligence state: persistent data surveillance, pattern-of-life analysis, and activity-based intelligence. These intelligence methods provide descriptive and/or predictive models of human behaviour that empower governments to generate intelligence about the actual and the potential subjects of counterterrorism investigations. The ethical implications of counterterrorism intelligence extend beyond the collection and use of data to the application of predictive modelling to dehumanised patterns of behaviour. This process has the potential to redefine the boundaries of the person, particularly by blurring the distinction between thoughts and actions which threaten the state.

## 1 Introduction

An international cohort of formal surveillance states emerged around the world during the early twenty-first century [84]. The group of countries that pursue national security through domestic surveillance and intelligence regimes grew throughout the 2000s into a wide and varied cohort in the 2010s [44, 85]. The political impetus to enhance national security from both internal and external threats was buttressed by the unprecedented availability and affordability of technology solutions to security

---

J. Hardy (✉)  
Rabdan Academy, Abu Dhabi, UAE  
e-mail: [jhardy@ra.ac.ae](mailto:jhardy@ra.ac.ae)

© The Author(s) 2021  
A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,  
Advanced Sciences and Technologies for Security Applications,  
[https://doi.org/10.1007/978-3-030-90221-6\\_7](https://doi.org/10.1007/978-3-030-90221-6_7)

105

challenges. The security state quickly became a surveillance state and now armed with the resources and reach into the everyday lives of its citizens, the security apparatus of many countries was transformed from a sprawling bureaucracy into an omnipresent institution. This spurred debates into the role of the state in safeguarding the liberty and privacy of individuals while also protecting the people and their interests from harm [47, 87]. While many of these debates remain unresolved, the rapid expansion of the state security apparatus, particularly in the advanced economies, has continued at increasing pace. Recent advancements in computer science and human understanding of analytical methods have empowered a modern intelligence state capable of deeper insight into the lives of individuals than the surveillance state that preceded it. Meanwhile, the post-9/11 political climate has enabled the rise of the intelligence state under the auspices of counterterrorism and national security [17].

The intelligence state combines the essential features of the security state and the surveillance state, protecting the homeland from internal and external threats through pervasive data collection and proactive policies aimed at counterterrorism and countering violent extremism [77]. It extends the reach of the security apparatus further by incorporating a broad range of analytical tools which enable analytical modelling and predictive analytics which stretch existing boundaries of surveillance [3]. Existing concerns about ethical conduct within mass surveillance programs are exacerbated by persistent data surveillance, pattern-of-life analysis, and activity-based intelligence. These analytical techniques provide both descriptive and predictive models of human behaviour that enable Governments unprecedented and invasive access to personal information for the purposes of enhancing counterterrorism. Routine encroachment of the intelligence state on the personal data of citizens has the potential to allow access to personal spaces by eroding the boundaries of privacy, to identify patterns of behaviour which fit risk profiles, and to create a system of control over access to information akin to Deleuze's [24] society of control.

The remainder of this chapter proceeds in three sections. The first section examines the gradual normalisation of mass surveillance as the state's technological omnipresence became a mechanism of discipline, security, and control. The second section examines the evolution of the surveillance state. It argues that the security state, which rose in tandem with Beck's [7] risk society, focusing more on technological solutions to domestic and international threats to homeland security than on traditional threats to national defence and security. The surveillance state rose in tandem with the evolution of big data. The generation, availability, and collection of an increasing number of data points from a rapidly expanding pool of sources led to the creation of modelling technologies that allowed states to build comprehensive profiles of citizens' patterns of life. The shift into an intelligence state has evolved alongside the fourth industrial revolution, utilising data surveillance, pattern analysis, machine learning, and predictive modelling to reduce the individual to a construct comprised of dividual data. The third section examines the 'dividual'<sup>1</sup> in the intelligence state and the implications of systematic analysis of dividual lives include the reduction of human lives to data points, behavioural analysis, and predictive algorithms. It

---

<sup>1</sup> 'Dividual' is a term coined by Gilles Deleuze, see discussion below.

argues that enhanced surveillance technologies routinely violate extant boundaries by giving the state access to previously non-observable thoughts and actions. This raises further questions about the rise of new technologies of control as innovations in Information and Communication Technologies (ICT) continue in the twenty-first century.

## 2 The Normalisation of Surveillance

Surveillance has been a function of the state in one form or another for much of recorded human history. The concept of an information state, which collected, processed, and stored information about its citizens has been popular among historians who generally link advances in technology as a defining characteristic of both modernity and the modernisation of the state [88]. The roles and functions of surveillance in the modern nation-state have served three key purposes. The first purpose was the ability to identify individuals in order to hold them accountable for aberrant or criminal behaviour in what Foucault [28] termed the disciplinary society. The second purpose was to manage the security and threat perceptions of individuals by promoting the state's ability to 'police' societies by deterring crime and deviance, to monitor society, and to prevent, mitigate, and respond effectively to major threats to security [16, 72]. The third purpose was to establish a system of control by enabling the state to construct public places, information systems, and individual interactions such that it facilitates the exchange of information and access to systems of information exchange [24]. This "panvasive" system is both pervasive and invasive, eroding the boundaries between the public and private, and collecting data from individuals indiscriminately [79].

### 2.1 *Surveillance as Control*

Foucault [28] depicted the emergence of societies of discipline as a product of successive societal institutions enclosing individuals into systems of rules and norms where behaviour was monitored, and compliance was rewarded and enforced. One key distinction between the societies of sovereignty of the past and the disciplinary societies from the eighteenth century onward was the purpose of governance. Where societies of sovereignty governed to retain power, raise taxes, and adjudicate over death, disciplinary societies sought to organise and administer both the individual and the collective [28]. Despite the popular conception of Bentham's panopticon as a foundation for contemporary theories of surveillance [60, 82], the act of surveillance is only one aspect of panopticism in the disciplinary society. Foucault's panopticon was an apparatus not only of pervasive surveillance and data collection, but also an environment characterised by ubiquitous institutional power, which could exert the political influence of potential surveillance over individuals at any time [25]. In

this conception, discipline is a mechanism of power that regulates the thoughts and behaviours of social actors through subtle means, increasing their docility and utility within society ([28], 231).

Deleuze [24] saw the disciplinary society as the foundation for a more comprehensive system of social compliance, once which was characterised not by surveillance and punishment, but by control. In a society of control, panopticism is not limited to the act of surveillance or the data being record. Rather, the panoptic system modulates social behaviour through access to information and opportunity ([25], 26). The metaphor of an omnipresent Big Brother perpetually watching society is less relevant to a society of control due to modes of participation in society which categorises and regulates the individual according to specified criteria which determine eligibility, inclusion, access, suspicion, and privilege ([49], 20). The society of control uses surveillance technologies to discriminate, assess, categorise, and profile individuals in what Gandy [30] has termed the “panoptic sort.” Deleuzian control presents a surveillance state which sustains a self-governing machine that exercises the subtle coercion of Foucault’s disciplinary society while constructing a societal landscape which moderates and modulates public places, information systems, and individual interactions. The society of control thereby facilitates both the exchange of information and individual access to collective systems of information exchange [25].

The society of control has been further enabled by advances in technology which fall into two categories. The first category is surveillance technologies which enhance the quality, volume, and integrity of collected data. The second category is analytical technologies which allow the state to construct new data from the products of its surveillance apparatus, to reduce risk, and to reduce the individual into a “bundle of data” to be collected, collated, and controlled ([23], 321).<sup>2</sup> Through a process of “surveillant assemblage,” the state is able to combine data it has collected and analysed about the individual and then generate an abstraction, or virtual identity, of that individual ([32], 608–610, [36], 11–12). Technological evolution throughout the early twenty-first century has changed the character, if not the nature, of the surveillance state with rapid advances in the means of controlling information and access. The trend toward mechanisms of control in the contemporary intelligence state has been supported by the increasing invisibility of surveillance in digital societies. This has been compounded by the growing acceptance of surveillance in the post-9/11 world and by the increasing complicity of individuals in subjecting themselves to data surveillance through consumer technologies [36, 63]. Increasing control and the normalisation of surveillance in many societies has brought the modern state to the precipice of existing ethical boundaries, sparking debates over the extent and appropriateness of mass surveillance programs justified under the auspices of counterterrorism and national security.

---

<sup>2</sup> Similar points were made by Daniel Solove in his 2004 *The Digital Person*, and by Adam Henschke in his 2017 *Ethics In An Age Of Surveillance*, where they talk about the use of surveillance technologies to create a ‘digital person’ or a ‘virtual identity’, respectively.



## 2.2 *Ethical Boundaries of the Surveillance State*

The notion of privacy in relation to surveillance is an evolving cultural phenomenon. Although the surveillance discourse often uses terminology that is associated with visual metaphors derived from historical societal experiences of state-based secret police, the dominant form of surveillance in the twenty first century is based on the collection and computation of data. Traditional metaphors of surveillance focus on the acts of watching and being watched. The state is conceived as a public eye surreptitiously watching both the public and private lives of citizens. Contemporary metaphors of surveillance focus more on mass collection of data points, creating vast stores of information which can be used to construct models of human behaviours, and an emphasis on empirical rather than visual terminology [2]. In contrast to traditional metaphors used to discuss surveillance, the contemporary discourse employs metaphors that are impersonal and formal. By focusing on technological concepts such as big data, metadata, and analytics, contemporary surveillance metaphors are reframing debates about privacy. Rather than discussing the boundaries of the state's presence in citizens' lives, contemporary metaphors draw attention to the pervasive technological capabilities of the state and a purported balance between security and liberty [35, 47].

One of the key arguments in debates about balancing security with liberty is the extent of the individual right to and expectation of privacy from observation by the state [37]. In this context, privacy can be framed as a general protection of the individual from unreasonable observation of behaviours in private spaces. In the twenty-first century the balance between the assumption of non-observable behaviour in private places, such as behind walls or in darkness, has been skewed by the pervasive collection of data by a range of actors, many of them non-state entities, including major transnational corporations. Online behaviours, activities, and identities have created a more complete digital personhood that allows the state to build more a complete picture of an individual this "virtual identity" has the potential to identify and expose an individual's private behaviours, attitudes, or beliefs [36, 80]. An exclusive focus on privacy for the sake of protecting individual privacy is reductionist, because privacy protections can be seen as safeguards for other aspects of citizen's lives, including restrictions on public expression and association, as well as protection from some forms of discrimination [9]. Information and data privacy can be seen as bulwarks against state encroachment into social and societal norms bounded and protected by contemporary liberal political values [70].

Information privacy is frequently conceived in terms of "fair information practices" which relate to communicative control ([49], 19). Communicative control is determined by the extent to which the subjects of data generation and collection know about and can influence how data about them are gathered, stored, analysed, shared, and used [49, 67]. This presumption of control extends beyond protecting personal details to include protection from institutional power. The power imbalance between surveillance apparatuses and the subjects of surveillance can be illustrated with three simple examples: blackmail, discrimination, persuasion [69]. The ability

to exploit personal data to create social sorting categories has expanded significantly in the post-9/11 era [44, 90]. Advancements in surveillance technologies have precepted an evolution of disciplinary societies into societies of control, empowering an omnipresent intelligence apparatus in relation to private citizens.

### 3 Technological Evolution of the Surveillance State

The contemporary national security state of the twentieth century was largely defined by great power struggles and deterrence. Throughout the post-Cold War period, the role of the state in providing security has undergone two significant changes. One change is the increased focus on internal threats to the security of the state under the banner of counterterrorism [68]. The other change is the impetus to monitor individuals at scale in order to detect indicators of threat behaviours [51]. This has intensified scrutiny of populations for the purpose of security, leading to the expansion of the surveillance functions of the state into an all-encompassing intelligence system capable of pervasive and invasive mass surveillance of foreign and domestic populations [78]. The evolution of the formal surveillance from a focus on homeland security to the modern intelligence state reflects the novel application of technology to security challenges and the coevolution of technology, data, and intelligence capability in the twenty-first century.

#### 3.1 *The Security State*

Security politics in the post-9/11 period have been dominated by counterterrorism in much the same way that perpetual threat of nuclear annihilation defined the Cold War period. The security state of the twenty-first century has been largely preoccupied by containing threats from non-state actors at home and abroad. Internal security, an enduring priority for the state, has become a central policy debate around the world. The contemporary expansion of the concept of security now includes a range of individual, transnational, and non-state issues [18]. Alongside a broad “new security agenda” [16], encompassing a non-traditional security threats, there has been increasing focus on domestic threats under the banner of homeland security [62]. Conceptual confusion about security, including what is, how it attained, and what it represents [6, 13], initially clouded attempts to create holistic homeland security policies [83].

States initially gravitated towards risk reduction and proactive security intelligence and law enforcement policies in order to meet the most prevalent threats to public safety. This mirrored a more general trend in policing by risk that had been developing in community policing for decades [53, 57]. However, security is both an objective measure of risk and safety and also a subjective interpretation of risk in a given situation [73]. Despite rapid advancements in risk reduction, the security

state quickly found itself lacking in managing risk perceptions. Public reactions to terrorist violence underscored the need to manage both risk reduction and public perceptions of safety. These converging impetuses led to widespread adoption of technological solutions to capability gaps in surveillance and intelligence.

Throughout the 2000s an increasingly capable and competent security state emerged. In some ways, the security state paralleled Ulrich Beck's [7] risk society. The overarching surveillance and intelligence apparatus of the security state served three main securitising functions in society. The first function was to restructure and repurpose institutions to include them in what would become the surveillance state. Burgeoning intelligence communities, intelligence fusion centres, and public-private data sharing arrangements [71] around the world suggest that the expansion of the security state has yet to reach its zenith. The second function was to build the capacity of the state to respond to domestic threats. Examples of the creation or enhancement of security intelligence, border protection, and law enforcement entities to bolster homeland security abound [62]. The third function was the subtle shift from policing the citizen actor within society to surveillance of the citizen threat to the state. Following the axiom that a person with nothing to hide has nothing to fear from surveillance [81], citizens have been effectively reframed as a potential source of threat to society. The security state thus paved the way for an expansionist surveillance state, justifying its penchant for panopticism under the guise of necessity.

### 3.2 *The Surveillance State*

The surveillance state emerged in the early twenty-first century alongside rapid advancements in ICT. New technologies have created a host of new opportunities and new vulnerabilities for intelligence collection and analysis, mostly related to the proliferation of big data. The surveillance implications of big data relate to both the unprecedented generation of data by individuals and a raft of new avenues for recording, accessing, and storing data [27, 36]. Major tech companies have benefited from nearly unilateral control over the capability to conduct pervasive data surveillance on individual consumers. This capability largely stems from mobile devices [43], but extends to a range of information services that major providers such as Amazon, Apple, and Google include in their product suites [54]. Governmental access to both the data and the collection platforms created by major tech companies and frequently used by consumers created a new conception of mass surveillance for the purposes of counterterrorism and homeland security [11].

From individual user profiles and search histories, to GPS location data and Bluetooth and Wi-Fi connections logs, to social media accounts and digital currency transactions, the sheer volume and speed of unique data generated by individuals engaging in their digital lives around the world is staggering [74]. The quantity, generation, and diversity of the data which governments collect, monitor, and analyse continue to increase rapidly. This is sometimes called the "five V's" of big data, which refers to the velocity, volume, value, variety, and veracity of data [29]. Access to these data

and evolving methods of analysis has created a variety of new forms of information that did not previously exist, such as comprehensive archives of location data collected by mobile devices ([36], 144–149). Meanwhile, the capacity to mine data sources, to monitor data flows in real-time, and to construct comprehensive models of individuals and their behaviour is a controversial issue because it enables a degree of passive surveillance that was not possible only a few decades ago ([36], 183–266). Clarke [21] termed this capability “data surveillance” and defined it as the ability to use data and analytics to effectively observe and record an individual, object, or organisation.

The technical collection of vast repositories of data by mobile devices has afforded the security apparatus of the state an unprecedented ability to gather and analyse data in ways which enable mass surveillance [75]. Meanwhile, new forms of data have coevolved with information and communication technologies, creating new methods of analysis and new kinds of intelligence [52, 65]. With the information collected through mass surveillance programs, intelligence analysts have become both better informed and burdened by the volume of data available to them. With expansive archives of digital information to sift through, the task of sorting, collating and categorising information has become more laborious [56]. The tasks involved in separating important details from trivial data and in deriving meaning from patterns and trends have become more intellectually challenging and increasingly resource intensive [27, 55]. One way of alleviating the burden of data collection on the state has been through the implementation of open source and crowd sourced data to complement data surveillance [61], which brings additional actors into contact with the mechanisms of the surveillance state.

### ***3.3 The Intelligence State***

The intelligence state is an evolution of the surveillance state that uses cutting edge technological and analytical capabilities to erode the previous boundaries of surveillance. The rise of the intelligence state has been enabled by the data generated through persistent invasive surveillance and empowered by analytical techniques such as network analysis, general Pattern-of-Life (POL) analysis, and Activity-Based Intelligence (ABI). The application of computer assisted modelling and predicative analysis to behavioural data creates new opportunities for the intelligence state to incorporate a broad range of technology-supported capabilities to conduct behavioural analysis, geospatial intelligence, POL analysis, and ABI [10]. The starting point for these methods is mass data collection and behavioural analytics. In general, behavioural analysis is a process of assessing and modelling routines, patterns, and events in interpersonal, public, and online behaviours in data about individuals and groups [48, 59]. Behavioural data can be enhanced with geospatial intelligence, which is derived from structured analysis of geographic, spatial, and imagery information [5, 38] and used to model physical, informational, and behavioural patterns.

Simple patterns in complex data can be highly revealing ([36], pp. 6–12). POL analysis can model patterns of association between people, places, and objects to identify nodes, events, patterns, and outliers in relational data [33]. These patterns in individual routines can provide significant insight into personal information without directly accessing private data or employing overt surveillance methods. Examples abound in contemporary society. One such example is the use of shopping data to personalise online advertisements on web sites and social media platforms. Another example is the use of mass gathered location data to monitor, predict, and manipulate traffic flows. A third example is the use of multiple data sources and methods of analysis to create an “ensemble effect,” which can illuminate personal preferences, behaviours, and patterns despite the limitations of any of the individual sources used [76]. Ensemble effects are sometimes used with crowd sourced and mass collected intelligence because they offer deeper and multifaceted insight into patterns and trends through data modelling [14].

ABI is a method of data modelling which focuses on actions and activities, incorporating contextual, biographical, and relational data which can be used to discover and systematise patterns and trends in a subject’s behaviours [22, 46]. ABI and similar predictive analytical methods empower the intelligence state to create models of subjects or targets of investigations, operations, and defensive countermeasures [10]. Intelligence models can be used for five basic purposes across the military, national security and law enforcement domains: description, collaboration, explanation, exploration and prediction [86]. Description is a method used to represent known details of an event, situation, or process. Collaboration allows teams of individuals to create a common representation of the modelled subject and then manipulate, update, and modify the shared model collectively. Explanation involves generating and testing hypotheses that potentially explain relationships between entity, event, or process data. Exploration is used to evaluate changes in the structure and dynamics of modelled subjects, explore causal influences between data, and anticipate behaviours. Prediction can be used to estimate likely events, to optimise processes and actions, and to pursue circumstances that are generally consistent with preferable outcomes [8, 86].

Activity-Based models have provided the intelligence state with new avenues for proactive and preventative actions to reduce risks, control crime, and identify individuals who display markers of targeted behaviours. For example, predictive policing models have enabled law enforcement agencies to increase resource allocation to locations and crime types deemed high risk, to develop intervention programs for specific crime types, and to identify functional, situational, and geographical factors for risks to safety and security [66]. Similarly, POL and ABI analyses have enabled intelligence and security bureaucracies to build models of adversary behavioural patterns and Modus Operandi, identify critical security events in progress, and enhance situational awareness [4]. These approaches to intelligence analysis require the data generated by the surveillance state as a fundamental input. However, the insight into the personal and private exceeds the boundaries that commonly exist

on physical, technical, and digital surveillance methods. For example, pattern detection algorithms used in threat behaviour models designed to support counterterrorism can bring individuals who have not been identified by the authorities under scrutiny. Where previously a warrant might have been needed to search a suspect's home or telephone records, their digital identity can be rapidly mined without facing traditional physical or temporal obstacles to surveillance.

## 4 The Dividual and the Intelligence State

The technological supremacy of the modern intelligence state over the individual citizen has the potential to reduce personhood to a sum of data points. Deleuze ([24], 5) coined the term 'dividual' to explain how a society of control could devolve the irreducible and autonomous agency of the individual to categories and classes with or without access. The dividual can also be conceived as a reducible unit of analysis, regarded by the intelligence state as combination of behavioural, biometric, communication, identity, location, and transaction data. The capability to exert control over society through moderating access to information, systems, and agency has grown immeasurably in the twenty first century. This raises two concerns for the evolving relationship between society and the intelligence state. The first concern is the emergent sense of self that is becoming more transparent both in public and private spaces. The second concern is the continuation of this trajectory in tandem with emerging technologies of control. The implications of these looming issues include the potential for a near-omniscient society of control and the gradual shaping of both civil society and citizenry into idealised state-designated models of behaviour and thought.

### 4.1 *The Transparent Self*

Contemporary societies are routinely subjected to levels and forms of surveillance which were not possible only a decade ago. The kinds of technologies that have been engrained in the daily routine of many people have also captured a detailed record about those routines ([69], 1936). For some, the persistence of surveillance technologies has been embraced as either a mixed blessing, permitting formerly unattainable levels of efficiency and security, or an acceptable cost for access to digital services ([49], 19). Certain forms of surveillance have been popularised in cinema and media [63], and have been accepted as a part of digital life in some societies. The kinds of surveillance technologies that people are increasingly conscious of in their everyday lives include communication metadata, GPS location data, social media feeds, and shopping activity. The kinds of surveillance that are less widely appreciated are user analytics applied to reading habits, browsing behaviours, and search term data [69].

These technologies have created a transparent self, one which is rendered more visible and less protected from external scrutiny through ubiquitous surveillance and data collection technologies [41]. Technology compromises the self in three distinct ways. The first form of compromise is the removal or reduction of boundaries, such as clothing and the human body itself, which is rendered transparent by non-invasive surveillance equipment such as millimetre wave, X-ray, and metal detecting scanners. The second form of compromise is the transparency of spaces formerly assumed to be non-observable, such as private spaces behind doors and walls or concealed by darkness, which are visible with electro-optical and remote audio sensing technologies. The third form of compromise is the blurring distinction between thoughts and actions which are deemed to pose a threat to the security of the state. The methods used to render internal thoughts observable are commonplace in the kinds of behavioural analytics used by commercial entities [1, 42]. Similar technologies are problematic for states where the line between criminal actions and thoughts about criminality is being eroded. One example of this in the counterterrorism domain is the state's focus on Countering Violent Extremism (CVE) by monitoring and intervening in ideological debates involving extremist content [31]. The distinction between having undesirable but legal thoughts and committing illegal behaviours is often blurred in the language used in CVE and counterterrorism policies [34, 89].

Gradual acceptance of the transparent self lends itself to the potential for a similarly gradual formation of a transparent society [12]. Such a society, defined by the transparency of its citizens and the degree of state control over access to networks, is a significant step closer to Deleuze's [24] society of control. The transparent society lays the foundation for the further encroachment of data generating technology into the personal lives and personal spaces of citizens. An example of this is the Social Credit Score [40] system used by the People's Republic of China (PRC) to incentivise and deter specific private behaviours in accordance with its preferences for citizen behaviour [45]. Since the introduction of the Social Credit Score system in 2007, the PRC has been able to rank, categorise, and sort its citizens, allocating resources and privileges to those who conform and denying access to those who do not [20]. This illustrates the extant and potential capability of the intelligence state to use surveillance, data and algorithms to exercise control over society [26].

## ***4.2 Emerging Technologies of Control***

The rapid expansion of technological solutions to security challenges in the post-9/11 era led to widespread public debate over the role of data surveillance in responding to the threat of domestic and transnational terrorism to the internal security of the state. During the 2000s, it was not feasible to use predictive analytics to effectively counter the threat of terrorism. The likelihood of errors, including both false positives and false negatives, was high due to limited data on the small-scale patterns of threat actor behaviour and the nascent capability of the security state to collect, store,

and analyse data [39]. This line of argument, which originated in the brief pre-smartphone period at the beginning of the twenty-first century, accurately reflected the limits of data availability, technical means of collection, analytical methods, and security intelligence models at the time. Much has changed since. While intelligence practitioners debate the relative effectiveness of emerging and enduring surveillance technologies [19, 50] there is little doubt that the ability of the intelligence state to leverage data analytics has improved dramatically [58, 64].

Although critics may contend that, in the words of William Burroughs [15], “control can never be a means to any practical end”, the modern intelligence state has created a system of control that serves as a means to greater levels of compliance and security than even societies of discipline could muster. The implications of near-certain future developments into surveillance and analytical technologies warrant further consideration. The implications of pattern analysis for the extension of control include the potential to create pervasive social monitoring and social sorting systems [49]. These systems would enhance the disciplinary power of predictive algorithms by enabling control over access to social benefits, information, knowledge systems, and opportunities [23]. A complex of disciplinary and control systems would be further empowered by progression in POL, ABI, and behavioural analytics [22]. Each of these technological advancements could permit further intrusion into the private and personal if not kept in check through robust protective measures. As such, the surveillance technologies currently under development by the intelligence state constitute technologies of control.

## 5 Conclusions

The rise of the modern intelligence state was, ostensibly, a technologically driven facet of the counterterrorism policies adopted by many countries in the post-9/11 era. Debates about the balance to be struck between privacy and security have largely focused on surveillance while the technical capability of the state evolved from a focus on surveillance to embracing intelligence and analytical modelling. Mirroring the conceptual evolution of disciplinary societies towards societies of control, the modern intelligence state has garnered an expansive reach into the previously personal and private spaces of citizens. This reach has been enabled by three emergent features of the modern intelligence state: persistent data surveillance, pattern-of-life analysis, and activity-based intelligence. Intelligence models provide insight into patterns of human behaviour and outlier activities which do not fit standard patterns. The dehumanising of data parallels a concomitant deindividualising of citizens, who may be controlled through mechanisms of social sorting. The nascent capacity to exercise influence by incentivising and deter specific behaviours is eroding the previous limits on the surveillance state. By monitoring and categorising citizen behaviours, the intelligence state does not just know the personal and private, it can to some extent shape what the private may be or think. With robust oversight,



these technologies may be beneficial to security in many ways. Nevertheless, the potential to redefine ethical boundaries requires strict attention.

## References

1. Abbot D (2014) *Applied predictive analytics: principles and techniques for the professional data analyst*. John Wiley & Sons, Hoboken, NJ
2. Agre PE (1994) Surveillance and capture: two models of privacy. *Inf Soc* 10(2):101–127
3. Andrejevic M (2012) Ubiquitous surveillance. In: Ball K, Haggerty KD, Lyon D (eds) *Routledge handbook of surveillance studies*. Routledge, London and New York, pp 91–98
4. Antony RT (2016) *Data fusion support to activity-based intelligence*. Artech House, Norwood, MA
5. Bacastow TS, Bellaifiore D (2009) Redefining geospatial intelligence. *Am Intell J* 27(1):38–40
6. Baldwin DA (1997) The concept of security. *Rev Int Stud* 23(1):5–26
7. Beck U (1992) *Risk society: towards a new modernity*. Sage, London
8. Moses LB, Chan J (2018) Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing Soc* 28(7):806–822
9. Bernal P (2016) Data gathering, surveillance and human rights: recasting the debate. *J Cyber Policy* 1(2):243–264
10. Biltgen P, Ryan S (2016) *Activity-based intelligence: principles and applications*. Artech House, Norwood, MA
11. Brayne S (2017) Big data surveillance: the case of policing. *Am Sociol Rev* 82(5):977–1008
12. Brin D (1999) *The transparent society: will technology force us to choose between privacy and freedom?* Basic Books, New York
13. Brooks DJ (2010) What is security: definition through knowledge categorization. *Secur J* 23(3):225–239
14. Bulger NJ (2016) The evolving role of intelligence: migrating from traditional competitive intelligence to integrated intelligence. *Int J Intell Secur Publ Affairs* 18(1):57–84
15. Burroughs WS (1959) *The naked lunch*. Olympia Press, Paris
16. Buzan B, Wæver O, de Wilde J (1998) *Security: a new framework for analysis*. Lynne Rienner, Boulder, CO
17. Byman D (2014) The intelligence war on terrorism. *Intell National Secur* 29(6):837–863
18. Caldwell D, Williams RE (2011) *Seeking security in an insecure world*. Rowman and Littlefield, Lanham, MA
19. Cayford M, Pieters W (2018) The effectiveness of surveillance technology: what intelligence officials are saying. *Inf Soc* 34(2):88–103
20. Chen Y, Cheung AS (2017) The transparent self under big data profiling: privacy and Chinese legislation on the social credit system. *Am J Comp Law* 12(2):356–377
21. Clarke R (1994) The digital persona and its application to data surveillance. *Inf Soc* 10(2):77–92
22. Craddock R, Watson R, Saunders W (2016) Generic pattern of life and behaviour analysis. In: 2016 IEEE international multi-disciplinary conference on cognitive methods in situation awareness and decision support (CogSIMA), San Diego, 21–25 March 2016
23. de Laat PB (2019) The disciplinary power of predictive algorithms: a Foucauldian perspective. *Ethics Inf Technol* 21(4):319–329
24. Deleuze G (1992) Postscript on the societies of control. *October* 59:3–7
25. Elmer G (2012) Panopticon—discipline—control. In: Ball K, Haggerty KD, Lyon D (eds) *Routledge handbook of surveillance studies*. Routledge, London and New York, pp 21–29
26. Erwin S (2015) Living by algorithm: smart surveillance and the society of control. *Humanities Technol Rev* 34:28–69
27. Ferguson AG (2017) *The rise of big data policing: surveillance, race, and the future of law enforcement*. New York University Press, New York

28. Foucault M (1995) *Discipline and punish: the birth of the prison*. Random House, New York
29. Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage* 35(2):137–144
30. Gandy O (1993) *The panoptic sort: a political economy of personal information*. Westview, Boulder CO
31. Gielen A-J (2019) Countering violent extremism: a realist review for assessing what works, for whom, in what circumstances, and how? *Terrorism Political Violence* 31(6):1149–1167
32. Haggerty K, Ericson R (2000) The surveillant assemblage. *Br J Sociol* 54(1):605–622
33. Hardy J, Lushenko P (2012) The high value of targeting: a conceptual model for using HVT against a networked enemy. *Def Stud* 12(3):413–433
34. Harris-Hogan S, Barrelle K, Zammit A (2016) What is countering violent extremism? Exploring CVE policy and practice in Australia. *Behav Sci Terrorism Political Aggression* 8(1):6–24
35. Hayden MV (2014) Balancing security and liberty: the challenge of sharing foreign signals intelligence. *Notre Dame J Law, Ethics Publ Policy* 19(1):247–260
36. Henschke A (2017) *Ethics in an age of surveillance: personal information and virtual identities*. Cambridge University Press, Cambridge
37. Henschke A (2020) Privacy, the internet of things and state surveillance: handling personal information within an inhuman system. *Moral Philosophy Politics* 7(1):123–149
38. Herchenrader T, Myhill-Jones S (2015) GIS supporting intelligence-led policing. *Police Pract Res* 16(2):136–147
39. Jonas J, Harper J (2006) *Effective counterterrorism and the limited role of predictive data mining*. CATO Institute, Washington, DC
40. Kobie N (2019) The complicated truth about China's social credit system. *Wired* 7 June
41. Lanzing M (2016) The transparent self. *Ethics Inf Technol* 18(1):9–16
42. Larose DT, Larose CD (2015) *Data mining and predictive analytics*. John Wiley & Sons, Hoboken, NJ
43. Laurila JK, Gatica-Perez D, Aad I, Blom J, Bornet O, Do T-M-T, Dousse O, Eberle J, Miettinen M (2012) The mobile data challenge: big data for mobile computing research. In: *Pervasive computing*, Newcastle
44. Levi M, Wall DS (2004) Technologies, security, and privacy in the post-9/11 European information society. *J Law Soc* 31(2):194–220
45. Liang F, Das V, Kostyuk N, Hussain MM (2018) Constructing a data-driven society: china's social credit system as a state surveillance infrastructure. *Policy Internet* 10(4):415–453
46. Long LA (2013) Activity based intelligence: understanding the unknown. *Intelligencier* 20(2):7–16
47. Lowe D (2016) Surveillance and international terrorism intelligence exchange: balancing the interests of national security and individual liberty. *Terrorism Political Violence* 28(4):653–673
48. Lyon D (2002) Everyday surveillance: personal data and social classifications. *Inf Commun Soc* 5(2):242–257
49. Lyon D (2003a) Surveillance as social sorting: computer codes and mobile bodies. In: Lyon D (ed) *Surveillance as social sorting: privacy, risk, and digital discrimination*, Routledge, London and New York, pp 13–30
50. Lyon D (2003b) Surveillance technology and surveillance society. In: Misa TJ, Brey P, Feenberg A (eds) *Modernity and technology*, The MIT Press, Cambridge, MA, pp 161–183
51. Lyon D (2014) Surveillance, snowden, and big data: capacities, consequences, critique. *Big Data Soc* 1(2):2053951714541861
52. Maguire M (2009) The birth of biometric security. *Anthropol Today* 25(2):9–14
53. Maguire M (2000) Policing by risks and targets: some dimensions and implications of intelligence-led crime control. *Polic Soc* 9(4):315–336
54. Mayer-Schönberger V, Ramge T (2018) A big choice for big tech: share data or suffer the consequences. *Foreign Aff* 97:48–54
55. McCue C (2015) *Data mining and predictive analysis: intelligence gathering and crime analysis*. Butterworth-Heinemann, Oxford

56. McCue C, Parker A (2003) Connecting the dots: data mining and predictive analytics in law enforcement and intelligence analysis. *Police Chief* 70(10):115–122
57. McGarrell EF, Freilich JD, Chermak S (2007) Intelligence-led policing as a framework for responding to terrorism. *J Contemp Crim Justice* 23(2):142–158
58. Meijer A, Wessels M (2019) Predictive policing: review of benefits and drawbacks. *Int J Publ Adm* 42(12):1031–1039
59. Mena J (2011) *Machine learning forensics for law enforcement, security, and intelligence*. CRC Press, Boca Raton
60. Miller J-A, Miller R (1987) Jeremy Bentham's panoptic device. *October* 41:3–29
61. Monahan T, Mokos JT (2013) Crowdsourcing Urban surveillance: the development of homeland security markets for environmental sensor networks. *Geoforum* 49:279–288
62. Mueller J, Stewart MG (2011) *Terror, security, and money: balancing the risks, benefits, and costs of homeland security*. Oxford University Press
63. Muir L (2012) Control space?: cinematic representations of surveillance space between discipline and control. *Urban Surveillance* 9(3):263–279
64. Nunn S (2001) Police technology in cities: changes and challenges. *Technol Soc* 23:11–27
65. Pell SK, Soghoian C (2014) Your secret stingray's no secret anymore: the vanishing government monopoly over cell phone surveillance and its impact on national security and consumer privacy. *Harvard J Law Technol* 28(1):2–75
66. Perry WL, McInnis B, Price CC, Smith S, Hollywood JS (2013) *Predictive policing: forecasting crime for law enforcement*. RAND Corporation, Santa Monica, CA
67. Pieters W (2011) The (social) construction of information security. *Inf Soc* 27(5):326–335
68. Richards J (2012) Intelligence dilemma? Contemporary counter-terrorism in a liberal democracy. *Intell National Secur* 27(5):761–780
69. Richards NM (2013) The dangers of surveillance. *Harv Law Rev* 126(7):1934–1965
70. Robbins S, Henschke A (2017) Designing for democracy: bulk data and authoritarianism. *Surveillance Soc* 15(3):582–589
71. Russel RL (2007) Achieving all-source fusion in the intelligence community. In: Johnson LK (ed) *Handbook of intelligence studies*. Routledge, London and New York, pp 189–198
72. Schneier B (2003) *Beyond fear: thinking sensibly about security in an uncertain world*. Copernicus Books, New York
73. Schneier B (2009) *Schneier on security*. Wiley Blackwell, Indianapolis
74. Schneier B (2015) *Data and goliath: the hidden battles to collect your data and control your world*. W. W. Norton, New York
75. Schuster S, van den Berg M, Larucea X, Slewe T, Ide-Kostic P (2017) Mass surveillance and technological policy options: improving security of private communications. *Comput Standard Interfaces* 50:76–82
76. Siegel E (2013) *Predictive analytics: the power to predict who will click, buy, lie, or die*. John Wiley & Sons, Hoboken, NJ
77. Sims J (2007) Intelligence to counter terror: the importance of all-source fusion. *Intell National Secur* 22(1):38–56
78. Slobogin C (2013) Panvasive surveillance, political process theory, and the non-delegation doctrine. *Georgetown Law Rev* 102:1721–1776
79. Slobogin C (2013) Rehnquist and panvasive searches. *Mississippi Law J* 82(2):307–328
80. Solove DJ (2004) *The digital person: technology and privacy in the information age*. New York University Press, New York
81. Solove DJ (2011) *Nothing to hide: the false trade-off between privacy and security*. Yale University Press, New Haven and London
82. Steadman P (2007) The contradictions of Jeremy Bentham's panopticon penitentiary. *J Bentham Stud* 9:1–31
83. Stolberg AG (2010) Making national security policy in the 21st century. In: Bartholomees JB (ed) *The U.S. army war college guide to national security issues*, Strategic Studies Institute, Carlisle, PA, pp 29–45

84. Svendsen A (2008) The globalization of intelligence since 9/11: frameworks and operational parameters. *Camb Rev Int Aff* 21(1):129–144
85. Tuinier P (2021) Explaining the depth and breadth of international intelligence cooperation: towards a comprehensive understanding. *Intell National Secur* 36(1):116–138
86. Waltz E (2014) *Quantitative intelligence analysis, applied analytic models, simulations, and games*. Rowman and Littlefield, Lanham, MA
87. Watt E (2017) The right to privacy and the future of mass surveillance. *Int J Human Rights* 21(7):773–799
88. Weller T (2012) The information state: an historical perspective. In: Ball K, Haggerty KD, Lyon D (eds) *Routledge handbook of surveillance studies*. Routledge, London and New York, pp 57–63
89. Williams MJ, Horgan JG, Evans WP (2016) The critical role of friends in networks for countering violent extremism: toward a theory of vicarious help-seeking. *Behav Sci Terrorism Political Aggression* 8(1):45–65
90. Zappalà G (2015) Killing by metadata: Europe and the surveillance-targeted killing nexus. *Global Affairs* 1(3):251–258

**John Hardy** is Assistant Dean of Graduate Studies at Rabdan Academy. John was previously Director of Security Studies at Macquarie University and a Research Fellow at the Australian National Security College (ANU). He has extensive experience in delivering academic and professional development training around the world, specialising in national security, intelligence, counterterrorism, and law enforcement. His current research projects focus on practical national security policy issues, such as the application of narrative analysis to extremist propaganda, mechanisms of coercive power in unequal political contests, and the use of emerging technologies to enhance security intelligence operations.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# “No Cracks, no Blind Spots, no Gaps”: Technologically-Enabled “Preventative” Counterterrorism and Mass Repression in Xinjiang, China



Michael Clarke

**Abstract** The Xinjiang Uyghur Autonomous Region (XUAR) of the People’s Republic of China (PRC) is now the site of the largest mass repression of an ethnic and/or religious minority in the world today. Researchers estimate that since 2016 over one million people (mostly ethnic Uyghurs) have been detained without trial in the XUAR in a system of “re-education” camps. Outside of the camps, the region’s Turkic Muslim population are subjected to a dense network of hi-tech surveillance systems, checkpoints, and interpersonal monitoring which severely limit all forms of personal freedom penetrating society to the granular level. This chapter argues that the erection of this “carceral state” has been propelled by a “preventative” counterterrorism that has incorporated key practices (e.g. greater reliance on new surveillance technologies) and discourses (e.g. Islamaphobia) of the “global war on terrorism” with the ideology of the Chinese Communist Party (CCP) in pursuit of the negation of the very possibility of “terrorism”. As such the contemporary situation in the XUAR represents not only the mass repression of an ethnic and religious minority by an authoritarian regime but also an example of the dystopian potentialities of ostensibly “neutral” technologies.

## 1 Introduction

Wrists and ankles strapped into a restraining “tiger chair”, a man is used as a subject with which to “train” artificial intelligence-assisted facial recognition technology to detect states of emotion. Minute changes in facial expression are analyzed by the facial recognition technology to determine whether the test subject possesses a “negative mindset” or a heightened state of anxiety, allegedly indicating a potential for anti-social behavior [82]. This is not a vision from a dystopic television series.

---

M. Clarke (✉)

Centre for Defence Research, Australian Defence College, Canberra, Australia  
e-mail: [michael.clarke@uts.edu.au](mailto:michael.clarke@uts.edu.au)

Australia–China Relations Institute, University of Technology Sydney, Sydney, Australia

© The Author(s) 2021

A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,  
Advanced Sciences and Technologies for Security Applications,  
[https://doi.org/10.1007/978-3-030-90221-6\\_8](https://doi.org/10.1007/978-3-030-90221-6_8)

121

On the contrary it is a lived reality in the Xinjiang Uyghur Autonomous Region (XUAR) in the far north-west of the People's Republic of China (PRC) where the Chinese state, in concert with a number of China's major surveillance technology companies, has striven to perfect new means of monitoring the region's Uyghur population. Researchers estimate that since 2016 over one million people (mostly ethnic Uyghurs) have been detained without trial in the XUAR in a system of "re-education" camps [22, 34]. Outside of the camps, the region's Turkic Muslim population are subjected to a dense network of hi-tech surveillance systems (including key elements of China's "social credit" system), checkpoints, and interpersonal monitoring which severely limit all forms of personal freedom penetrating society to the granular level [62, 95]. The objective, as XUAR Chinese Communist Party (CCP) deputy leader Zhu Hailun asserted in 2017, is to ensure that there are "no cracks, no blind spots, no gaps" in the state's surveillance of the region [87].

This chapter argues that the erection of this "carceral state" has been propelled by a "preventative" counterterrorism that has incorporated key practices (e.g. greater reliance on new surveillance technologies) and discourses (e.g. Islamophobia) of the "global war on terrorism" with the ideology of the Chinese Communist Party (CCP) in pursuit of the negation of the very possibility of "terrorism". As such the contemporary situation in the XUAR represents not only the mass repression of an ethnic and religious minority by an authoritarian regime (although it is most certainly that) but also an example of the dystopian potentialities of ostensibly "neutral" technologies.

In this latter respect, I argue that it has been the intersection of technologically-enabled surveillance with the CCP's evolving ideological concept of "social management" that defines the practice and effects of China's "preventative" counterterrorism in the XUAR. Descriptions of the system of surveillance erected in the XUAR as simply the manifestation of a new type of "police state" only capture part of the story. Control of the region's Uyghur population is but one objective of the CCP in XUAR. Indeed, as Richard Jenkins reminds us, surveillance is but "a means to an end", namely the "protection" and "management" of either the population-at-large or specific segments thereof ([54]: 162). The case of Chinese counterterrorism in the XUAR reveals the Chinese state's propensity to be much more explicit in its desire—relative to governments in the liberal West—to pursue the active (and often coercive) 'management' of specific segments of its population.

China's counterterrorism policy is in fact highly suggestive of processes of "high modernism" described by James C. Scott in which the state seeks to legitimate the "rational design of social order" ([78]: 4) through the centralization, collection, and processing of information. Scott suggested that the imposition of such "high modernism" tended to correlate with crises (e.g. economic depression, social revolution or war) and authoritarianism. In particular, the manner in which the system of pervasive surveillance intersects with the CCP's practices of ideological "re-education" in XUAR demonstrates how surveillance—from the state's perspective—serves goals beyond mere control of subject populations by identifying, categorizing, and ascribing sanction to individuals to produce "transformed" citizens. That surveillance is a central enabler of the CCP's social engineering objective is demonstrated by the assertion of a XUAR government spokesman on 25 May, 2021, that heightened

security measures and “re-education” were required in order to “remove extremist thoughts” from Uyghur minds and “transform” them from “ghosts” into “humans” [77]. While the ‘threat of terrorism and religious extremism’ has stimulated the development of ‘new forms of centralized surveillance, monitoring and identification’ regardless of regime type ([84]: 61) the Chinese state has thus been able to instrumentalize the threat of Uyghur ‘terrorism’ and ‘religious extremism’ to further a deeper end—the remoulding an entire population’s behaviours in the name of cultural assimilation. As we will see, this has significant human rights implications and poses a challenge for liberal democracies who espouse the importance of values like free movement, individual privacy, and free speech.

## 2 Chinese Colonialism and Uyghur ‘Terrorism’ in Xinjiang

Despite China’s contemporary claim that Xinjiang (literally ‘new dominion’ or ‘new frontier’) has been ‘an inseparable part of the unitary multi-ethnic Chinese nation’ since the Han Dynasty (206 BC—24 AD), it often remained beyond Chinese dominion due to its geopolitical position as a ‘Eurasian crossroad’ and the ethno-cultural dominance of Turkic and Mongol peoples [15].

After experiencing significant autonomy from the Republic of China (1911–1949), Xinjiang was “peacefully liberated” by the People’s Liberation Army (PLA) in October 1949 and the CCP confronted the question of “how to run an empire without looking like colonialists” [65]. Their answer—recognition of the region’s 12 non-Han *minzu* (nationality or ethnic group) and implementation of a system of “national regional autonomy”—in theory, was meant to ensure that “beneath the supreme central CCP power” the various *minzu* were to stand as equals, their individual culture, language and practice of religion respected and protected [65]. In practice, however, this was accompanied by tight political, social and cultural control, encouragement of Han Chinese settlement, and state-led economic development, backed by the repression of overt manifestations of opposition and dissent by the security forces ([7]: 120–129).

After the collapse of the neighbouring Soviet Union in 1991, the focus of Beijing’s concerns regarding the security of Xinjiang shifted from state-based threats to largely non-state ones driven by the convergence of the Islamic revival in neighbouring Central Asia and Afghanistan and relative weakness of the post-Soviet states [5].

Under Deng’s successor, Jiang Zemin, the question of Xinjiang’s economic development assumed national importance under his Great Western Development (GWD) campaign, formally launched in 2000. Under the GWD Xinjiang was envisaged as becoming an industrial and agricultural base and a trade and energy corridor for the national economy. Central to the state’s developmental agenda was a focus on a variety of “mega-projects” such as massive oil and natural-gas pipelines and infrastructure developments linking Xinjiang with Central and South Asia and the various sub-regions of Xinjiang with each other and the interior of China [1].

While bringing economic development, such projects also created a variety of new socio-economic pressures—encouragement of further Han settlement, rapid urbanization, and environmental degradation—that exacerbated interethnic tensions [11]. However many Uyghurs felt they had not benefitted from economic development due to a variety of factors including: the concentration of Xinjiang’s urban centers and industry in the north of the province; targeting of state investment in large infrastructure projects in which companies have tended to employ Han Chinese; and widening rural–urban disparities [4, 14, 17].

This period not coincidentally saw an appreciable increase in Uyghur unrest and militancy. Data collected by the University of Maryland’s Global Terrorism Database, for example, records 135 attacks in Xinjiang across the 1992 and 2017 period resulting in 767 fatalities [39]. However, those figures count as terrorist attacks a number of incidents—such as the 7 July 2009 violence in Xinjiang’s capital, Urumqi, which resulted in 184 fatalities—even though they’re more accurately defined as inter-ethnic rioting or communal violence prompted by the long-term marginalisation of the Uyghur population (see [9, 16, 19]). Omitting this incident alone decreases the death toll from terrorism in Xinjiang to 583 over the 25-year period.

This only increased in intensity after the events of 9/11 as the Party-state instrumentalized the threat and discourse of “global terrorism” to justify and expand its efforts to monitor and control key markers of Uyghur identity such as religious observance/piety. It is clear that 9/11 provided Beijing with the stimulus to reframe its efforts in Xinjiang as ‘counterterrorist’ rather than simply counter ‘separatist’ in nature. This began immediately after 9/11, when Beijing released its first documentation of terrorist incidents in Xinjiang, blaming a previously unknown group, the ‘East Turkestan Islamic Movement’ (ETIM), for ‘over’ 200 ‘terrorist incidents’ between 1990 and 2001 [50]. A number of high-profile attacks in more recent years, such as the October 2013 SUV attack in Tiananmen Square and the April 2014 Kunming railway station mass stabbing attack, reinforced China’s official narrative that it faces a genuine terrorist threat stemming from Xinjiang ([28], 73–74).

The presence of the al-Qaeda-aligned TIP in Syria from 2012 onward was important in assisting Beijing in its desire to paint Uyghur militancy as intimately interconnected with global ‘jihadist’ forces [28]. Despite these linkages, however, there is in fact little available evidence of TIP’s direct involvement in attacks in Xinjiang. TIP has claimed responsibility for a number of high-profile attacks, such as the so-called SUV attack of October 2013 in Tiananmen Square, but, Jacob Zenn notes ‘only a 2011 hit-and-run attack in Kashgar’ has been ‘credibly proven’ to have been organised by the group from Afghanistan [92]. Chinese state media however leveraged the presence of Uyghurs in Syria to argue that Beijing’s hard-line in Xinjiang was warranted. The English-language tabloid, *Global Times*, for instance, published an editorial on 12 August 2018 asserting that China’s hard-line approach in the region had prevented it from becoming ‘China’s Libya’ or ‘China’s Syria’ [40]. Prior to institution of China’s hard-line, it continued, ‘young people were brainwashed by extremist thoughts and manipulated by terrorist organizations’, resulting in terrorist attacks not only in Xinjiang but also ‘in places such as Tiananmen Square of Beijing and Kunming Railway Station’ [40].



### 3 China’s Counterterrorism Policy: Toward ‘Enduring Peace’

It was in this period of heightened official concern with the threat of terrorism to Xinjiang that China’s form of “preventative” counterterrorism took shape. Of particular note was the development of a new strategy based on the integration of traditional Maoist ‘mass line’ (*qunzhong luxian*) mobilization and social control with new forms of technologically-enabled surveillance and policing. The “mass line”—in which the CCP sought to organize and mobilize the Chinese population in support of the Party’s objectives and policies through regular mass campaigns—was a regular and defining feature of Chinese governance under Mao Zedong’s leadership. The aim, as Elizabeth Perry ([68], 33) has noted, “was to prevent bureaucratic inertia by recruiting grassroots enthusiasts to augment (and in some cases override) local party and government cadres so as to advance the central leaders’ agendas”. Indeed, the “mass line” was a chief means through which the Maoist project of the “achievement of Utopian social goals by means of class struggle and the cleansing of society in order to create an egalitarian society” was enacted ([2], 324). For Xi, the return of mass campaigns is a necessary measure to “standardize party procedures, curb corruption and enhance the party’s overall competence” and thus ensure not only the sustainability of one-party rule but also the country’s “great national rejuvenation” [46].

In Xinjiang, this has entailed intensified Party-state interventions in society in order to ensure the twin goals of “stability” and “development”. The need for such intervention in the CCP’s estimation was underlined by inter-ethnic violence in the region’s capital of Urumqi on 5 July 2009 (referred to in China as the 7/5 Incident) The 7/5 Incident, in which officially 194 people were killed over two days of inter-ethnic violence, convinced influential leaders that the twin strategies of “national regional autonomy” and state-led economic development upon which Chinese governance had rested since 1978 had exacerbated rather than assuaged long-standing sources of disgruntlement with the Chinese state.

The CCP’s immediate response to the 7/5 Incident was focused on replacement of senior party figures in Xinjiang (including Urumqi CCP secretary, Li Zhi, and long-serving Xinjiang CCP chairman, Wang Lequan), deployment of People’s Armed Police and Special Police Units to XUAR, and a renewed focus on “stability maintenance” and economic development [72]. In this latter regard, then President Hu Jintao, at the first Central Xinjiang Work Forum (XJWFI) of the CCP held 17–19 May 2010 unveiled a “Xinjiang support package” including targeted central government investment and infrastructure spending. The objective, according to Hu, was to achieve “leapfrog development” of the region that would lift it’s GDP to the national average by 2015 and thus contribute to “ethnic unity” and “social stability” [55, 91].

The new XUAR CCP chairman, Zhang Chunxian, thus embarked on what was dubbed a ‘two handed’ policy in the region of both ‘hard’ and ‘soft’ measures focused on ‘stability maintenance’ work and improving ‘people’s livelihood’ that would consolidate the Party’s ‘grassroots infrastructure’ throughout the region [63].

After 2009 official public security spending increased rapidly, with much of was spent on the introduction of high-definition surveillance cameras across public spaces in Xinjiang, including in mosques [30]. By the next year at least 40,000 high-definition “Eagle-Eye” surveillance cameras equipped with ‘riot-proof’ casings were fitted on buses, in schools, and in shopping centres, as well as on the streets of urban areas to increase police presence in key places, vital sectors and public areas ([37, 85]: 58).

Zhang’s era of ‘two-handed’ policy however was also accompanied by a transition in how the CCP conceived of the relationship between development, identity and security. For much of the post-Mao era the Party’s strategy in Xinjiang had rested on the assumption that development would resolve its “Uyghur question” by breaking down the traditional cultural, religious and social ties that underpinned Uyghur identity and thus secure the region. After 7/5, however, economic development per se was viewed as no longer sufficient. Rather, the question now was what obstacles prevented development from achieving the goal of integration and what should the Party do about it. An answer emerged from the debates about a so-called “second generation” of ethnic minority policy after 2009 [60]. Party-affiliated scholars such as Ma Rong, Hu Angang, and Hu Lianhe argued that the “first generation” of policy—based on ethnic equality and “national regional autonomy”—had solidified ethnic boundaries, ethnic elites, and notions of “separateness” [61]. The direction of ethnic minority policy since has demonstrated that their conclusion has been that there is something intrinsic to Uyghur identity that blocks the path to the Party’s vision of modernization, and hence, integration. Advocates of “second generation” policy therefore argued that ethnic policy must discard the nominal pluralism and preferential policies of the past in favour of an approach that explicitly sought the “mingling”, “fusing” or “standardization” of ethnic groups with a supra-national conception of the Chinese “state-nation” (*zhongguo minzu*) [61].

In March 2012 however Xi Jinping (then Vice-President) rebuked Zhang Chunxian’s ‘two handed’ approach. Xi noted not only that ‘Xinjiang work’ held a ‘particularly important strategic position in the overall work of the party and the state’ but that XUAR CCP officials must ‘unswervingly insist on *both* development and stability’ and ‘hold high the banner of unity’ [81]. In October 2014, the National Security Commission (NSC) also established a National Anti-Terrorism Intelligence Centre to strengthen anti-terrorism intelligence gathering in order to boost its counterterrorism pre-emptive and preventive capabilities ([53]: 190). The State Ethnic Affairs Commission (SEAC), which had previously led the development and implementation of governance of ethnic minority regions, was also down-graded as the locus of ethnic minority governance after 2009 as provincial level CCP United Front Work Department ‘offices assumed primary responsibility for ethnic work in ethnic minority regions, with SEAC officials left to follow the direct lead of their Party counterparts’ ([21]: 491).

The need for greater Party control over ethnic minority governance was underlined by a number of violent incidents in or connected to Xinjiang in 2013 and 2014 including the so-called ‘SUV attack’ in Tiananmen Square on 28 October 2013 and the Kunming Railway station attack of 1 March 2014 that officials blamed on ‘radicalized’ Uyghurs [8, 73]. These incidents contributed to Xi’s decision after a

Politburo meeting on 19 December 2013 that the CCP would abandon Zhang’s ‘two-handed’ policy in Xinjiang. State media reported that the Party’s ‘prime task’ in Xinjiang would now be the pursuit of ‘social stability and an enduring peace’ [90]. This approach subsequently took on the language of counter terrorism in response to the perceived extremist ideologies found in the region.

As would become clear over the following two years, the goal of ‘enduring peace’ in Xinjiang would be sought through reinvigoration of Maoist ‘mass line’ forms of Party mobilization, implementation of new forms of technological surveillance and intensive ‘de-extremification’ work, including ‘concentrated re-education training’ of those deemed to be at risk of ‘extremism’. After further violence in May and July 2014, Zhang Chunxian voiced the starkest rhetoric yet exhorting a meeting of the XUAR Party Committee to fight a ‘people’s war against terrorism’ that would not only ‘cut weeds’ but also ‘dig out the roots’ of extremism [52]. This resulted in accelerated arrests and trials of suspected ‘terrorists’—including public, mass sentencing rallies of Uyghur suspects—and ongoing sweeps of Uyghur neighborhoods and mosques in search of potential militants and their weapons [58].

These trends of increased technological surveillance combined with ideological ‘re-education’ of those defined as potential ‘extremists’ were accelerated in 2016 under the new XUAR CCP Chairman Chen Quanguo. Chen had in fact implemented a policing system of ‘grid style management’ during his previous role as Party leader in Tibet (2011–2015) that segmented ‘urban communities into geometric zones’ policed by ‘convenience’ police stations connected to CCTV cameras and police databases enabling greater surveillance capabilities [94]. In Xinjiang, Chen implemented ‘grid management’ and integrated it with the CCTV surveillance systems established under his immediate predecessor, resulting in a multi-tiered policing system based on exponential recruitment of contract police officers to man ‘convenience’ police stations [93, 95]. Additional surveillance measures—including compulsory fitting of GPS trackers in motor vehicles, use of facial recognition scanners at checkpoints and major public amenities and installation of ‘nanny apps’ that wipe smartphones of so-called “subversive” material—were also implemented under Chen’s watch [32, 70]. The purpose of such a system was explicitly detailed by Chen in a speech on 18 August 2017 in which he gave instructions for the “party, government, military, police, soldiers and civilians” of XUAR to implement “comprehensive, round-the-clock and three dimensional prevention control” in order to “deny *any* opportunity to hostile forces and violent terrorists” to undermine the region’s “stability” ([88]. Emphasis added).

#### **4 Seeing Like the CCP: ‘Social Management’, Counterterrorism and ‘Re-Education’**

The methodology that has been central to the pursuit of this ‘comprehensive, round-the-clock and three dimensional prevention control’ has been the concept of ‘social

management'. Samantha Hoffman notes that 'social management' embodies an effort to optimise 'interactions vertically (within the Party), horizontally (between agencies), and holistically, between the Party and society' in order 'to improve governance capacity to shape, manage, and respond to social demands' [48]. It ultimately seeks to enhance the 'legibility' of citizens and to make them pliable subjects to be engineered and thus controlled by the state [78]. As James C. Scott reminds us, the 'utopian, immanent, and continually frustrated goal of the modern state' has been 'to reduce the chaotic, disorderly, constantly changing social reality beneath it to something more closely resembling the administrative grid of its observations' thereby rendering citizens and the spaces in which they inhabit more transparent to the gaze of the state legible and thus responsive to central manipulation and control (Ibid). In fact the 'security state' erected in Xinjiang under the tenures of XUAR CCP Party chiefs Zhang Chunxian and Chen Quanguo has enabled the Party-state to undertake 'social sorting' on a large scale. 'Social sorting', in Jenkins' conception, seeks the 'identification and ordering of individuals in order to "put them in their place" within local, national and global "institutional orders"', and to thus ascribe to them particular penalties, constraints or sanctions according to their categorization ([54]: 160). As will be detailed below, this is what has occurred to large numbers of Xinjiang's Uyghur population. The ends to which such means are deployed is not simply to increase the Party-state's ability to 'see' the Uyghur population in all its permutations but also to manufacture the consent of Uyghur population and enable it to actively mould and shape those individuals into 'productive' and pliable citizens.

The CCP's project of making of Uyghurs 'legible' has been highlighted in its recommitment to expand the security presence throughout the region, particularly through the use of enhanced surveillance capabilities, and by means of the legalization and institutionalization of ideological and political 'thought' work on its citizens. It has now been well-documented that technological innovation has been vital to this project with the use of facial recognition and iris scanners at checkpoints, train stations and gas stations, collection of biometric data for passports, and mandatory apps to cleanse smartphones of subversive material now fact of everyday life for the Uyghur population [33, 66, 67]. The data collected is then aggregated by an app used by security personnel, the Integrated Joint Operations Platform (IJOP), to report "on activities or circumstances deemed suspicious" and to prompt "investigations of people the system flags as problematic" [49].

A closer examination of the legislative and discursive architecture that has been built around the surveillance apparatus reveals how precisely the CCP decides who is problematic or "untrustworthy". Legislatively, there have been a number of shifts at the national and provincial level here. First, in December 2015 the National People's Congress (NPC) passed China's first national "anti-terrorism" law, providing an expansive and ambiguous definition of terrorism that further enables the state to criminalise a wide array of actions. The law states that terrorism is:

Any advocacy or activity that, by means of violence, sabotage, or threat, aims to create social panic, undermine public safety, infringe on personal and property rights, or coerce a state organ or an international organization, in order to achieve political, ideological, or other objectives [86].

Second, the XUAR government’s March 2017 ‘de-extremification’ regulations revealed the state’s objective to categorize and punish those it defines as ‘deviant’ and ‘abnormal’. These regulations defined ‘extremification’ not only as ‘speech and actions under the influence of extremism, that imbue radical religious ideology, and reject and interfere with normal production and livelihood’ but also explicitly identified fifteen ‘primary expressions’ of ‘extremist thinking’, such as ‘wearing, or compelling others to wear, gowns with face coverings, or to bear symbols of extremification’, ‘spreading religious fanaticism through irregular beards or name selection’, and ‘failing to perform the legal formalities in marrying or divorcing by religious methods’ [31]. This list was subsequently expanded to include another sixty signs of “extremism” including such behaviors as quitting smoking and men growing long beards. This, Joanne Smith-Finley argues, amounted to a criminalization of ‘all religious behaviours, not just violent ones’, leading ‘to highly intrusive forms of religious policing’ that violate and humiliate Uyghurs [80]. ‘Extremism’, in the CCP’s definition, is thus conflated with everyday markers and practices of the Uyghur profession of Islam.

Third, China’s White Paper of 16 August 2019 on ‘Vocational Education and Training in Xinjiang’ [51], neatly demonstrates the way in which surveillance is not simply about control but also the production of particular socio-political outcomes. In this instance, surveillance has enabled the CCP to define and regulate Uyghur values, beliefs, and loyalties in such a way as to ensure individuals become ‘useful’ subjects for maintaining the regime’s political security [57]. While defining ‘terrorism and extremism’ as ‘common enemies of human society’ and Xinjiang as the ‘main battlefield of China’s fight against terrorism and de-extremization’, the document asserted that the state must not only deal with ‘terrorist crimes in accordance with the law’ but also ‘educate and rescue personnel infected with religious extremism and minor crimes’ in order to treat ‘both symptoms and the root causes’ of religious extremism. The document asserts that it is through ‘education and training’ that Xinjiang will ‘achieve social stability and enduring peace’ by promoting development and increasing people’s overall income [51].

However it is a 52 gigabyte internal police dataset from the Urumqi Public Security Bureau (PSB) in the capital of the region, obtained by *The Intercept* [41] and analysed in detail by Darren Byler [3], that perhaps best demonstrates the intersection of surveillance technology and the ideological underpinnings of the current repression in Xinjiang. Beginning in 2013 the Urumqi PSB began experimenting with mobile scanning devices that ‘integrated 3G mobile technology through smart phone terminals and VPN-enabled database synchronization in order to allow rapid individual identity authentication’ ([3], 11). By 2017 this had been upgraded to allow police in the capital to scan and read ID cards, ‘instantly linking ID numbers, issuers, and photos’ of the individual being checked to the IJOP. These ‘social incident reports’—some 250 million rows of data in the files obtained by *The Intercept*—list the date and time of the encounter, the precinct, name, ID number, gender, ethnicity and phone number of the suspect. They describe the reason why the individual was flagged and if they warrant further investigation. They also list the geolocation of the encounter’ (Ibid, 12). The Urumqi PSB used this system to primarily monitor the capital’s

Uyghur and Kazakh population, subjecting them to regular checks, ‘targeted observation’, household searches, monitoring of familial and community relationships and mosque attendance (Ibid, 12–13).

Yet such ‘technology systems cannot simply be plugged in and work their magic on their own’ but ‘are only as good as the data they are trained on’ (Ibid, 19–20). Here, the data from the Urumqi PSB files demonstrates that the authorities have trained the technology to identify and aggregate actionable intelligence based on ideologically-defined criteria. Significantly, the Urumqi PSB’s hand-held mobile scanning devices are also armed with a ‘digital forensics’ tool called the ‘Anti-Terrorism Sword’ that can scan ‘smartphones and other electronic devices in less than two minutes, attempting to match materials to a base dataset of as many as 53,000 flagged audio, video, picture and text files that had been deemed related to religious extremism or terrorism’ [26]. The ‘Anti-Terrorism Sword’ also enables the police to access ‘private social media, email and instant messaging applications to assess the phone owner’s digital history and social network’ (Ibid). Through such means, as recounted to Darren Byler by an ethnic Kazakh police officer, the PSB was able to ascertain whether or not a person ‘had worn an Islamic veil, had installed WhatsApp or had traveled to Kazakhstan’ (Ibid). All of these data points—from an individual’s record of mosque attendance through to social media use or travel history—are used to flag an individual for further investigation or detention [43].

Thus the monitoring of everyday life in Uyghur neighbourhoods is geared to identify and respond to what the CCP has defined as key markers of ideological deviancy. From government officials describing Uyghur “extremism and terrorism” as a “tumour” to the equation of religious observance to an “illness”, the CCP’s discourse frames central elements of Uyghur identity as pathologies to be “cured” [35]. That such pathologizing of Uyghur identity guides official policy was made plain by a CCP Youth League official’s justification of “re-education” in October 2017. “Being infected by religious extremism and violent terrorist ideology”, the official asserted, “is like being infected by a disease that has not been treated in time, or like taking toxic drugs” and even after completing the “re-education process” individuals “must remain vigilant, empower themselves with the correct knowledge, strengthen their ideological studies ... to bolster their immune system against the influence of religious extremism and violent terrorism, and safeguard themselves from being infected once again” [71]. This frames the Uyghur population as a “virtual biological threat to the body of society” [74]. The ultimate “cure” for this biopolitical threat posed by Uyghur identity, as stated in an internal CCP document of March 2018, is to “break their lineage, break their roots, break their connections, and break their origins” [44]. As the dataset from the Urumqi PSB demonstrates, however, it is the surveillance apparatus erected in the region that enables the ‘social sorting’ that is central to the operation of the ‘re-education’ system in Xinjiang.

## 5 Conclusion

The CCP, as demonstrated above, has actively sought to manage and reshape the behaviours of the Uyghur population through a security and surveillance apparatus that makes them “transparent” to the gaze of the state and hence eminently controllable. This, as two theorists at the Xinjiang Police University argued in 2016, amounted to the emergence of a “Xinjiang model” of counterterrorism that would combine what they defined as the “war model” of counter-insurgency adopted by the US military in Iraq and Afghanistan with China’s own “public security model” and “governance model” [83]. The “public security model” was built on “the construction of the anti-terrorism intelligence system”—embodied in “grid management” and technological surveillance initiatives noted above—which would provide security forces with “the ability to obtain information on signs, tendencies ... related to violence and terrorism” and thereby enhance “social prevention and control capabilities” [83]. The “governance model”, in turn, focuses on the long-term “resolution of ethnic and religious ideological issues” that give rise to “extremism” and “terrorism”. Here, Wang and Shan asserted that as religious “extremism” is an “ideological” problem it must be solved “by ideological methods” ([83]: 25). This entailed sustained “education” of the population in order to “reject the brainwashing of distorted religious views” and thereby increase their “immunity to extreme terrorism” (Ibid).

However we must recognize, as Wang and Shan’s exposition of a “Xinjiang model” of counterterrorism indicates, that the CCP’s implementation of a surveillance-enabled form of what it terms counterterrorism has not taken place in a vacuum. Rather, it is part of a globalization of “countering violent extremism” (CVE) strategies and discourses that aim to both reduce “extremism” with non-military instruments and sanctions available (or created) under domestic law and/or to prevent such “extremism” from occurring in the first place through interventions at the individual and societal level to ameliorate “root causes” of such behaviors. Simultaneously, the case of Xinjiang also reveals the unique aspects of the CCP’s practice of this globalized mania for preventative forms of counterterrorism. For the Xinjiang Police University theorists Wang and Shan the central objective of the “Xinjiang model” is to undermine what the CCP sees as a root cause of terrorism in Xinjiang: religion. “Extreme religion”, Wang and Shan assert, “attempts to change the true face of national culture and block exchanges and fusion among all ethnic groups” and as such the Party’s “cultural guidance” must assist “people of all ethnic groups” to “move closer to secularization and modernization”. The central implication is that “there must be an acceleration of ‘the deep fusion’ of Chinese culture in Xinjiang” in order to eliminate terrorism [25]. As we have seen, the CCP’s ability to break the connection between markers of Uyghur religious and cultural identity and what it perceives as “extremism” has been fundamentally enabled by both the implementation of new forms of surveillance, and reinvigoration of older forms (e.g. Maoist “mass line”), that permit “social sorting” on a mass scale. Here, the “Xinjiang model” of counterterrorism emerges as nothing less than a



new instrument with which to secure China's colonial project in Xinjiang to control, exploit and "remake" the region and its Turkic Muslim peoples.

The evolution of the "Xinjiang model" of counterterrorism has broad implications not only for the trajectory of the CCP's governance of the PRC but also for global dynamics of surveillance technologies. With respect to the governance of Xinjiang and the PRC, the system erected in Xinjiang fixes Uyghurs (and other Turkic Muslim minorities) in place, makes them "transparent" to the gaze of the state and hence eminently controllable. The technologies that have permitted the Party-state to monitor and control the population in Xinjiang potentially sets China on the path to becoming a 'responsive tyranny', in which digital technologies empower the state to act pre-emptively and to identify and quash opposition in advance, on the basis of clues gleaned from its many channels of mass information collection ([18], 64).

This technologically-enabled system of surveillance and control also intersects with global dynamics in a number of key ways. First, the utilisation of specific technological innovations such as DNA sequencing, metadata analysis, facial recognition technology and machine learning are becoming increasingly deployed by states throughout the globe across the both the global North and global South in the name of public safety and, especially, counterterrorism [10, 38, 59]. This trend makes it both easier for the Chinese state to construct a justificatory narrative around its system of control and for the state's various security apparatuses and bureaucracies to engage with and learn from international partners. Second, the Chinese state's engagement with, and prioritisation of, surveillance technologies has resulted in the increased direct involvement of a number of Chinese and global tech companies in provision of both technology and components to the "security state" in Xinjiang [29, 64]. Finally, President Xi Jinping's multi-billion dollar Belt and Road Initiative (BRI) is intended to invest only in physical infrastructures—but also in the infrastructure and technology necessary to create a "digital Silk Road." Much of this investment is coming from China's major tech companies, including Alibaba, Huawei, and ZTE [56]. In addition, the manner in which China's tech companies seem to be investing so heavily in emerging surveillance technology suggests that its gaze is broad: It wants to address Beijing's surveillance imperatives at home but also secure customers abroad [36]. While such companies are undoubtedly more focused on profit it is also likely that the "presence of Chinese engineers, managers, and diplomats will reinforce a tendency among developing countries, especially those with authoritarian governments" to adopt China's approach of ensuring that technology serves the interests of a homogeneous state [79].

While the system of pervasive surveillance—both of the 'mass line' and technologically-enabled varieties—combined with the practices of "re-education" in XUAR arguably represents an extreme example of the deeply dystopic potentialities of such "high modernist" ideologies and technologies of social control, the spread and potential normalisation of such a 'surveillance-industrial' complex through appeals to 'counter-terrorism' imperatives constitutes an emerging global challenge to norms of basic human rights that must be guarded against.



## References

1. Becquelin N (2004) Staged development in Xinjiang. *China Q* 2004:358–378
2. Brown K, Bērziņa-Čerenkov U (2018) Ideology in the era of Xi Jinping. *Chin J Political Sci* 23(3):323–339
3. Byler D (2021) Chinese infrastructures of population management on the new silk road. In: Denmark A (ed) *Essays on the rise of China and its implications*. Wilson Center, Washington DC, pp 7–34
4. Chaudhuri D (2011) Minority economy in Xinjiang—a source of Uyghur resentment. *China Report* 46(1):9–27
5. Clarke ME (2008) China’s “war on terrorism” in Xinjiang: human security and the causes of violent Uighur separatism. *Terrorism Political Violence* 20(2):271–301
6. Clarke M (2010) Widening the net: China’s anti-terror laws and human rights in the Xinjiang Uyghur autonomous region. *Int J Human Rights* 14(4):542–558
7. Clarke M (2011) *Xinjiang and China’s rise in central Asia—a history*. Routledge, London
8. Clarke M (2015) China and the Uyghurs: the “Palestinization” of Xinjiang? *Middle East Policy* 22(3):127–146
9. Cliff T (2012) The partnership of stability in Xinjiang: state–society interactions following the July 2009 unrest. *China J* 68:79–105
10. Condell J et al (2018) Automatic gait recognition and its potential role in counterterrorism. *Stud Conflict Terrorism* 41(2):151–168
11. Karrar H (2017) Resistance to state-orchestrated modernization in Xinjiang: the genesis of unrest in the multiethnic frontier. *China Inf* 32(2):183–202
12. Kim H (2004) *Holy war in China: the Muslim rebellion and Khanate in Chinese Central Asia, 1864–1877*. Stanford University Press, Stanford
13. Li Y (2018) *China’s assistance program in Xinjiang: a sociological analysis*. Rowman & Littlefield, Lanham
14. Liu AH, Peters K (2017) The hanification of Xinjiang, China: the economic effects of the great leap west. *Stud Ethn Natl* 17(2):265–280
15. Millward J (2007) *Eurasian crossroads: a history of Xinjiang*. Columbia University Press, NY
16. Millward J (2009) Does the 2009 Urumchi violence mark a turning point? *Cent Asian Surv* 28(4):347–369
17. Piazza JA et al (2018) Digging the “ethnic violence in China” database: the effects of inter-ethnic inequality and natural resources exploitation in Xinjiang. *China Rev* 18(2):121–154
18. Qiang X (2019) The road to digital unfreedom: president Xi’s surveillance state. *J Democr* 30(1):53–67
19. Ryono A, Galway M (2015) Xinjiang under China: reflections on the multiple dimensions of the 2009 Urumqi uprising. *Asian Ethnicity* 16(2):235–255
20. Song T et al (2019) Policy mobilities and the China model: pairing aid policy in Xinjiang. *Sustainability* 11(13):3496
21. Zhao T, Leibold J (2020) Ethnic governance under Xi Jinping: the centrality of the united front work department and its implications. *J Contemp China* 29(124):487–502
22. Batke J (2019) Where did the one million figure for detentions in Xinjiang’s camps come from? *China file*, 8 January. <https://www.chinafile.com/reporting-opinion/features/where-did-one-million-figure-detentions-xinjiangs-camps-come>
23. Boxun xinwen (2014) ‘公安部长郭声琨年内第三次赴新疆调研反恐’, 6 August 2014. <https://www.boxun.com/news/gb/china/2014/08/201408062003.shtml>
24. Byler D (2017) Imagining re-engineered Muslims in northwest China. *Milestones: Commentary on the Islamic World*, 20 April. <https://www.milestonesjournal.net/photo-essays/2017/4/20/imagining-re-engineered-muslims-in-northwest-china>
25. Byler D (2019) Preventative policing as community detention in northwest China. *Made Chin J* 25 October. <https://madeinchinajournal.com/2019/10/25/preventative-policing-as-community-detention-in-northwest-china/>

26. Byler D (2020) The Xinjiang data police. *Noema Magazine* 8 October. <https://www.noemamag.com/the-xinjiang-data-police/>
27. Clarke M (2016) Uyghur militants in Syria: the Turkish connection. *Terrorism Monitor* 14(3). [http://www.jamestown.org/programs/tm/single/?tx\\_ttnews%5Btt\\_news%5D=45067&tx\\_ttnews%5BbackPid%5D=26&cHash=4c5a2b3135329c40a8c418a3ad2966c8#.VsJrGHluli4](http://www.jamestown.org/programs/tm/single/?tx_ttnews%5Btt_news%5D=45067&tx_ttnews%5BbackPid%5D=26&cHash=4c5a2b3135329c40a8c418a3ad2966c8#.VsJrGHluli4)
28. Clarke M (2020) Uyghur militancy and terrorism: the evolution of a ‘Global’ Jihad? In: Smith T, Schulze K (eds) *Exporting global jihad vol. 2: critical perspectives from Asia and North America*, I. B. Tauris, London, pp 73–98
29. Chen S-CJ (2018) SenseTime: the faces behind China’s artificial intelligence unicorn. *Forbes* 7 March. <https://www.forbes.com/sites/shuchingjeanchen/2018/03/07/the-faces-behind-chinas-omniscient-video-surveillance-technology/#63549f474afc>
30. China Daily (2010) Xinjiang security funding increased by 90%, 13 January. [http://www.chinadaily.com.cn/china/2010-01/13/content\\_9311035.htm.v](http://www.chinadaily.com.cn/china/2010-01/13/content_9311035.htm.v)
31. China Law Translate (2017) Xinjiang Uyghur autonomous region regulation on de-extremification’, 30 March. <https://www.chinalawtranslate.com/en/xinjiang-uyghur-autonomous-region-regulation-on-de-extremification/>
32. Coca N (2018) China’s Xinjiang surveillance is the dystopian future nobody wants. *Engadget* 22 February. <https://www.engadget.com/2018/02/22/china-xinjiang-surveillance-tech-spread/>
33. Daly A (2019) Algorithmic oppression with Chinese characteristics: AI against Xinjiang’s Uyghurs’. [https://strathprints.strath.ac.uk/71586/1/Daly\\_GISW2019\\_Algorithmic\\_oppression\\_Chinese\\_characteristics\\_AI\\_against\\_Xinjiang\\_Uyghurs.pdf](https://strathprints.strath.ac.uk/71586/1/Daly_GISW2019_Algorithmic_oppression_Chinese_characteristics_AI_against_Xinjiang_Uyghurs.pdf)
34. de Hahn P (2019) More than 1 million Muslims are detained in China—but how did we get that number? *Quartz*, 5 July. <https://qz.com/1599393/how-researcherestimate-1-million-uyghurs-are-detained-in-xinjiang/>
35. Dooley B (2018a) ‘Eradicate the tumours’: Chinese civilians drive Xinjiang crackdown. *Yahoo News*, 26 April. <https://www.yahoo.com/news/eradicate-tumours-chinese-civilians-drive-xinjiang-crackdown-051356550.html>
36. Dooley B (2018b) Chinese firm’s cash in on Xinjiang’s growing police state. *Yahoo News*, 27 June. <https://au.news.yahoo.com/chinese-firms-cash-xinjiangs-growing-police-state-033907491--spt.html>
37. Famularo J (2018) “Fighting the enemy with fists and daggers”: the Chinese communist party’s counter-terrorism policy in the Xinjiang Uyghur autonomous region. In: Clarke M. (ed) *Terrorism and counter-terrorism in China: domestic and foreign policy dimensions*, Oxford University Press, NY
38. Ganor B (2019) Artificial or human: a new era of counterterrorism intelligence? *Stud Conflict Terrorism* 1–20. <https://doi.org/10.1080/1057610X.2019.1568815>
39. Global Terrorism Database (2020) Xinjiang keyword search. In: National consortium for the study of terrorism and responses to terrorism (START), University of Maryland. <https://www.start.umd.edu/gtd/search/Results.aspx?search=Xinjiang&sa.x=48&sa.y=4>
40. Global Times (2018) Protecting peace, stability is top of human rights agenda for Xinjiang, 12&nbsp;August. <http://www.globaltimes.cn/content/1115022.shtml>
41. Grauer Y (2021) Revealed: mass Chinese police database. *Intercept*, 29 January. <https://theintercept.com/2021/01/29/china-uyghur-muslim-surveillance-police/>
42. Greitens SC et al (2019, 2020) Counterterrorism and preventive repression: China’s changing strategy in Xinjiang. *Int Secur* 44(3):9–47
43. Greer T (2018) 48 ways to get sent to a Chinese concentration camp. *Foreign Policy*, 13 September. <https://foreignpolicy.com/2018/09/13/48-ways-to-get-sent-to-a-chinese-concentration-camp/>
44. Government Information Public Platform of Kashi (2018) Notice on printing and distributing the “responsibility plan for the key points of inspection work in Kashgar region in 2018”, 6 March. <http://kashi.gov.cn/Government/PublicInfoShow.aspx?ID=2851>
45. Hayes B (2012) The surveillance-industrial complex. In: Ball K et al (eds) *Routledge handbook of surveillance studies*, Routledge, London, pp 167–175

46. Heath T (2013) ‘Xi’s mass line campaign: realigning party politics to new realities. *China Brief* 13(16). <https://jamestown.org/program/xis-mass-line-campaign-realigning-party-politics-to-new-realities/>
47. Hierman B (2007) The pacification of Xinjiang: Uighur protest and the Chinese state, 1988–2002. *Prob Post-Communism* 54(3):48–62
48. Hoffman S (2017) Managing the state: social credit, surveillance and the CCP’s plan for China. *China Brief* 17(11). <https://jamestown.org/program/managing-the-state-social-credit-surveillance-and-the-ccps-plan-for-china/>
49. Human Rights Watch (2019) China’s algorithms of repression, 1 May. <https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance>
50. Information Office of the State Council of the PRC (2002) East Turkistan terrorist forces cannot get away with impunity, 21 January. <http://www.china.org.cn/english/2002/Jan/25582.htm>
51. Information Office of the State Council of the PRC (2019) Full text: vocational education and training in Xinjiang. *Xinhua*. [http://www.xinhuanet.com/english/2019-08/16/c\\_138313359.htm](http://www.xinhuanet.com/english/2019-08/16/c_138313359.htm)
52. Jacobs A (2014) China says nearly 100 killed in week of unrest in Xinjiang. *New York Times*, 3 August. [https://www.nytimes.com/2014/08/04/world/asia/china-says-nearly-100-are-killed-in-week-of-unrest-in-xinjiang.html?\\_r=5](https://www.nytimes.com/2014/08/04/world/asia/china-says-nearly-100-are-killed-in-week-of-unrest-in-xinjiang.html?_r=5)
53. Ji Y (2016) China’s national security commission: theory, evolution and operations. *J Contemp China* 25(98)
54. Jenkins R (2012) Identity, surveillance and modernity: sorting out who’s who’. In: Ball K et al (eds) *Routledge handbook of surveillance studies*, Routledge, London
55. Jia C, Zhu Z (2010) Xinjiang support package unveiled. *China Daily*, 21 May. [https://www.chinadaily.com.cn/china/2010-05/21/content\\_9874981.htm](https://www.chinadaily.com.cn/china/2010-05/21/content_9874981.htm)
56. Joplin T (2018) China’s global surveillance-industrial complex. *Al Bawaba*, 21 June. <https://www.albawaba.com/news/long-form-china%E2%80%99s-global-surveillance-industrial-complex-1141152>
57. Klimeš O (2018) Advancing “ethnic unity” and “de-extremization”: ideational governance in Xinjiang under “new circumstances” (2012–2017). *J Chin Political Sci* 23(3)
58. Koplowitz H (2014) China Uighur conflict: gang knife attack in Xinjiang blamed on Islamic terrorists. *Int Bus Times*, 29 July. <https://www.ibtimes.com/china-uighur-conflict-gang-knife-attack-xinjiang-province-blamed-islamic-terrorists-1642368>
59. Lehr P (2018) *Counter-terrorism technologies: a critical assessment*. Springer, Cham, Switzerland
60. Leibold J (2013) Ethnic policy in China: is reform inevitable? *Policy Stud* 68, East-West Center, Honolulu. <https://scholarspace.manoa.hawaii.edu/bitstream/10125/30617/ps068.pdf>
61. Leibold J (2018) Hu the uniter and the radical turn in China’s Xinjiang policy. *China Brief* 18(16). <https://jamestown.org/program/hu-the-uniter-hu-lianhe-and-the-radical-turn-in-chinas-xinjiang-policy/>
62. Leibold J (2019) The spectre of insecurity: the CCP’s mass internment strategy in Xinjiang. *China Leadership Monitor*, 1 March. <https://www.prcleader.org/leibold>
63. Lin, M. (2014). ‘Winning Uyghurs’ hearts’, *Global Times*, 11 May, <http://www.globaltimes.cn/content/859697.shtml>
64. Lin L, Chin J (2019) U.S. tech companies prop up China’s vast surveillance network. *Wall Street J*, 26 November, <https://www.wsj.com/articles/u-s-tech-companies-prop-up-chinas-vast-surveillance-network-11574786846>
65. Millward J (2019) *Reeducating Xinjiang’s Muslims*. *New York Review of Books*. <https://www.nybooks.com/articles/2019/02/07/reeducating-xinjiangs-muslims/>
66. Ma A (2019) China uses an intrusive surveillance app to track its Muslim minority. *Bus Insider*, 11 May. <https://www.businessinsider.com.au/how-ijop-works-china-surveillance-app-for-muslim-uyghurs-2019-5?r=US&IR=T>
67. Mozur P (2019) One month, 500,000 face scans: how China is using A.I. to profile a minority. *New York Times*, 14 April. <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>

68. Perry EJ (2011) From mass campaigns to managed campaigns: “constructing a new socialist countryside”. In: Heilman S, Perry EJ (eds) *Mao’s invisible hand: the political foundations of adaptive governance in China*, Brill, Lieden, pp 30–61
69. Qiu Y (2014) Turkey’s ambiguous policies help terrorists join IS jihadist group: analyst. *Global Times*
70. Radio Free Asia (2017) Vehicles to get compulsory GPS tracking in Xinjiang, 20 February. <https://www.rfa.org/english/news/uyghur/xinjiang-gps-02202017145155.html>
71. Radio Free Asia (2018) Xinjiang political “re-education camps” treat Uyghurs “infected by religious extremism”: CCP youth league, 8 August. <https://www.rfa.org/english/news/uyghur/infected-08082018173807.html>
72. Ramzy A (2010) A year after Xinjiang riots, ethnic tensions remain. *Time*, 5 July. <http://content.time.com/time/world/article/0,8599,2001311,00.html>
73. Roberts S (2013) Tiananmen crash: terrorism or cry of desperation? *CNN*, 31 October. <https://edition.cnn.com/2013/10/31/opinion/china-tiananmen-uyghurs/index.html>
74. Roberts S (2018) The biopolitics of China’s ‘war on terror’ and the exclusion of the Uyghurs. *Crit Asian Stud* 50(2)
75. Roberts S (2020) *The war on the uyghurs: China’s campaign against Xinjiang’s Muslims*. Manchester University Press
76. Rosenblatt N (2016) All jihad is local: what ISIS’ files tell us about its fighters. *New America Foundation*, Washington DC
77. rthk.hk (2021) Xinjiang camps ‘turning ghosts into humans’, 25 May, <https://news.rthk.hk/rthk/en/component/k2/1592581-20210525.htm>
78. Scott JC (1997) *Seeing like a state: how certain schemes to improve the human condition have failed*. Princeton University Press, Princeton, NJ
79. Segal A (2018) When China rules the web: technology in service of the state. *Foreign Affairs*. [https://www.foreignaffairs.com/articles/china/2018-08-13/when-china-rules-web?cid=nlc-fa\\_fatoday-20180814](https://www.foreignaffairs.com/articles/china/2018-08-13/when-china-rules-web?cid=nlc-fa_fatoday-20180814)
80. Smith Finley J (2018) Islam in Xinjiang: ‘De-extremification’ or violation of religious space?, *Asia Dialogue*, 15 June, <https://theasiadialogue.com/2018/06/15/islam-in-xinjiang-deextremification-or-violation-of-religious-space/>
81. Song J (2012) 习近平参加新疆团审议 强调坚持稳定压倒一切, [Xi Jinping participates in Xinjiang delegation review], *Zhongguo xinwen wang*, 12 March, reprinted in <https://news.qq.com/a/20120312/000209.htm>
82. Wakefield J (2021) AI emotion-detection software tested on Uyghurs, *BBC News*, 26 May, <https://www.bbc.com/news/technology-57101248>
83. Wang D, Shan D (2016) 反恐研究与新疆模式 [Studies on anti-terrorism and the Xinjiang mode]. *情报杂志 [J Intell]* 35(11):20–26
84. Weller T (2012) *The information state: a historical perspective on surveillance*. In: Ball K, Hegarty K, Lyon D (eds) *Routledge handbook of surveillance studies* (Routledge, London). pp 57–63
85. Wu Q (2014) Urban grid management and police state in China: a brief overview. *China Change*, 12 August. <https://chinachange.org/2013/08/08/the-urban-grid-management-and-police-state-in-china-a-brief-overview/>
86. Xinhua (2015) China adopts first counter-terrorism law in history, 27 December. [http://www.xinhuanet.com/mil/2015-12/28/c\\_128574674.htm](http://www.xinhuanet.com/mil/2015-12/28/c_128574674.htm)
87. Xinjiang R (2017a) Zhu Hailun zai Akesu diqu zhaokai jiceng ganbu zuotanhui’ [Zhu Hailun convenes a grassroot cadre forum in the Aksu region], 20 April. <http://news.sohu.com/20170420/n489632907.shtml>
88. Xinjiang R (2017b) [Chen Quanguo gave instructions on doing the current stability work in Xinjiang: build a copper wall and iron wall to fight terrorism and maintain stability to ensure the overall harmony and stability of Xinjiang], *Renming wang*. Reprinted in <http://xj.people.com.cn/n2/2017/0819/c186332-30628706.html>
89. Xinjiang XW (2015) 新疆伊宁县: 开展 ‘去极端化’ 集中教育 [Yining county, Xinjiang: carrying out “de-radicalization” intensive education], 12 January. [http://www.agri.cn/DFV20/XJ/dfzx/dfyw/201501/t20150113\\_4331764.htm](http://www.agri.cn/DFV20/XJ/dfzx/dfyw/201501/t20150113_4331764.htm)

90. Yang J (2014) Xinjiang to see “major strategy shift”. Global Times, 9 January. <http://www.globaltimes.cn/content/836495.shtml#.UtS1ivaFZ0Q>
91. Yue H (2010) Hand in hand: China unveils a partner assistance program to propel Xinjiang toward economic prosperity and social stability. Beijing Rev 23. [http://www.bjreview.com.cn/Cover\\_Story\\_Series\\_2010/2010-06/07/content\\_277589.htm](http://www.bjreview.com.cn/Cover_Story_Series_2010/2010-06/07/content_277589.htm)
92. Zenn J (2011) Jihad in China? Marketing the Turkistan Islamic party. Terrorism Monit 9(11), <https://jamestown.org/wp-content/uploads/2011/03/TM0090>
93. Zenz A, Leibold J (2017a) Xinjiang’s rapidly evolving security state. China Brief, 14 March. <https://jamestown.org/program/xinjiangs-rapidly-evolving-security-state/>
94. Zenz A, Leibold J (2017b) Chen Quanguo: the strongman behind Beijing’s securitization strategy in Tibet and Xinjiang. China Brief 17(12). <https://jamestown.org/program/chen-quanguo-the-strongman-behind-beijings-securitization-strategy-in-tibet-and-xinjiang/>
95. Zenz A, Leibold J (2019) Securitizing Xinjiang: police recruitment, informal policing and ethnic minority co-optation. China Q 1–25. <https://doi.org/10.1017/S0305741019000778>

**Michael Clarke** is a Senior Fellow at the Centre for Defence Research, Australian Defence College, Canberra, and a Visiting Fellow at the Australia–China Relations Institute, University of Technology Sydney. His major areas of research concern the history and politics of the Xinjiang Uyghur Autonomous Region, People’s Republic of China (PRC) and Chinese foreign and security policy. He is the author of *Xinjiang and China’s Rise in Central Asia—A History* (Routledge 2011); editor (with Anna Hayes) of *Inside Xinjiang: Analysing Space, Place and Power in China’s Muslim North–West*, (Routledge 2016), editor of *Terrorism and Counterterrorism in China: Domestic and International Dimensions*, (Oxford University Press 2018) and editor of *The Xinjiang Emergency: Exploring the Causes and Consequences of China’s Mass Repression of Uyghurs*, (Manchester: Manchester University Press, February 2022). His commentary has been published by *Foreign Policy*, *War on the Rocks*, *The Wall Street Journal*, *CNN*, *The Diplomat*, *South China Morning Post* and *The National Interest* amongst others.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# Privacy, Encryption and Counter-Terrorism



Seumas Miller and Terry Bossomaier

**Abstract** Privacy is an important moral right but so is security, including security from terrorist attacks. Encryption protects privacy rights but also affords protection to terrorists and impedes legitimate counter-terrorist operations. This chapter analyses this ethical dilemma.

It is agreed on all sides that there is an important right to privacy, but that security is also important and, in particular, security from terrorist attacks. However, security requirements dictate that privacy rights be infringed at times, e.g. in the case of intercepting emails or phone conversations between terrorists. Moreover, encryption is obviously a good thing since it protects privacy, but potentially problematic if it unreasonably impedes legitimate counter-terrorism operations. The ethical dilemma in this area is exemplified by the following two relatively recent events.

Firstly, there was the conflict between Apple and the FBI [7, 22]. In December 2015, Syed Farook killed 14 people in San Bernardino [4, 26]. The FBI suspected that his phone may have contained information which could implicate others involved in the planning of the attack, or in possible future attacks. However, an Apple iPhone allows only 10 attempts to unlock the phone via its four-digit password before the phone is wiped. Apple refused the FBI request to remove the 10 attempts limit. Ultimately Apple did not have to back down, since a third party succeeded in cracking the phone (including, conceivably, by bypassing or shutting down the auto-erase feature by some means).

Secondly, in mid-2020, Operation Venetic in the UK and coordinated operations in Europe made news when very large criminal networks in the UK and in Europe were destroyed as a result of access to their supposedly secure EncroChat mobile

---

S. Miller (✉) · T. Bossomaier  
Charles Sturt University, Canberra, Australia  
e-mail: [semiller@csu.edu.au](mailto:semiller@csu.edu.au)

S. Miller  
TU Delft, Delft, Netherlands

University of Oxford, Oxford, England

phones. Joseph Cox in a thorough article on Vice Motherboard reported that in the Netherlands alone, “the investigation has so far led to the arrest of more than 100 suspects, the seizure of drugs (more than 8000 kilo cocaine and 1200 kilo crystal meth), the dismantling of 19 synthetic drugs labs, the seizure of dozens of (automatic) fire weapons, expensive watches and 25 cars, including vehicles with hidden compartments, and almost EUR 20 million in cash” [5]. In the UK, over 700 arrests—including of crime bosses—have been made, and two tons of drugs (worth over £100 million) have been seized [28]. The phone, which was basically a customised Android phone, provided end-to-end encryption, i.e. email, text messages and voice calls are encrypted on the phone and not decrypted until they reach the destination phone. It is thought the phone was not decrypted but rather hacked into, since malware was apparently found on the EncroChat device itself, meaning that it could potentially read the messages written and stored on the device before they were encrypted and sent over the internet (see Sect. 2). While Operation Venetic concerned criminal organisations primarily engaged in drug dealing, money-laundering, weapons distribution and murder of rival criminals, phones with end-to-end encryption (see Sect. 2) are known to be widely used by terrorists, thus this law enforcement achievement is highly germane to counter-terrorism operations.

These two events graphically illustrate the importance of encryption in law enforcement and in counter-terrorism, in particular. On the one hand, encryption provides privacy protection to ordinary citizens, confidentiality protection to legitimate businesses and, for that matter, confidentiality to police and other security agencies engaged in crime-fighting and counter-terrorism. On the other hand, encryption also affords protection to drug cartels, human traffickers and, of particular interest here, terrorist organizations.

To address this ethical question, we undertake three main tasks. Firstly, we offer an analysis of the nature and moral significance of privacy, including its relationship to confidentiality, autonomy and security, in the context of the counter-terrorism responses of liberal democratic states. Secondly, we provide a description of relevant cryptographic technologies. One focus here will be on WhatsApp, an open architecture, in the sense of being described in a white paper,<sup>1</sup> but not meeting the open-source criterion discussed below, for which we can describe key exchange structure. We explain how the keys work, with minimal mathematics, and the challenges they present to security agencies. By describing the technical issues in some detail, we show how it is that high level end-to-end encryption is, in effect, invulnerable to decryption but also how devices that use such encryption are, nevertheless, vulnerable by virtue of their use of passwords and the possibility of being hacked and the insertion of malware. This section is of particular importance, given the central role this technology has come to play in terrorism and counter-terrorism and given, also, the lack of understanding of the actual powers and limitations of this technology due to its highly technical nature. Our third main task in this article is to provide a discussion of the privacy rights and security needs in relation to encryption in the overall context of the counter-terrorism policies of liberal democratic states.

---

<sup>1</sup> <https://www.whatsapp.com/security>.



## 1 Privacy/Confidentiality, Autonomy and Security

The notion of privacy has proven difficult to adequately explicate [6, 9, 12, 16, 18, 24, 31, 34]. Nevertheless, there are a number of general points that can be made. First, privacy is a right that people have in relation to other persons and organisations with respect to: (a) the possession of personal information about themselves by other persons and by organisations, e.g. data stored in telecommunication company, technology company, or government databases, (b) the observation/perceiving of themselves—including of their movements, relationships and so on—by other persons, e.g. via CCTV or mapping metadata to determine geolocation history; (c) the interception of their communications, e.g. phone conversations, emails.

Second, the right to privacy is closely related to the more fundamental moral value of autonomy. Roughly speaking, the notion of privacy delimits an informational and observational ‘space’ i.e. the private sphere. However, the right to autonomy consists of a right to decide what to think and do and, of relevance here, the right to control the private sphere and, therefore, to decide *whom to exclude and whom not to exclude* from it. So, the right to privacy consists of the right to exclude organisations and other individuals (the right to autonomy) both from personal information and facial images, and from observation and monitoring (the private sphere). Naturally, the right to privacy is not absolute; it can be overridden. Moreover, its precise boundaries are unclear; a person does not have a right not to be casually observed in a public space but, arguably, has a right not to have their movements tracked via their smartphone, albeit this right can be overridden under certain circumstances, e.g. if they are terrorism suspects.

Third, a degree of privacy is necessary simply in order for people to pursue their personal projects, whatever those projects might be. For one thing, reflection is necessary for planning, and reflection requires a degree of freedom from the distracting intrusions, including intrusive surveillance, of others. For another, knowledge of someone else’s plans can lead to those plans being thwarted (e.g. if one’s political rivals can track one’s movements and interactions then they can come to know one’s plans in advance of their implementation), or otherwise compromised, (e.g. if who citizens vote for is not protected by a secret ballot, including a prohibition on cameras in private voting booths, then democracy can be compromised). *Autonomy*—including the exercise of autonomy in the public sphere—requires a measure of privacy.

Thus far we have described privacy and autonomy, considered as the rights of a *single* individual. However, it is important to consider the implications of the infringement, indeed violation, of the privacy and autonomy rights of the whole citizenry by the state (and/or other powerful institutional actors, such as corporations). Such violations on a large scale can lead to a power imbalance between the state and the citizenry and, thereby, undermine liberal democracy itself. The surveillance system imposed on the Uighurs in China, incorporating a full range of technologies including phone metadata, facial recognition, DNA, etc., graphically illustrates the



risks attached to large scale violations of privacy and related autonomy rights if governments use them in a discriminatory manner [8, 11, 17, 20].

In light of the above analysis of privacy, and especially its close relationship to autonomy, we are entitled to conclude that some form of privacy is a constitutive human good. As such, infringements of privacy ought to be avoided. That said, as mentioned above, privacy can reasonably be overridden by security considerations under some circumstances, such as when lives are at risk. After all, the right to life is, in general, a weightier moral right than the right to privacy.

Individual privacy is sometimes confused with anonymity, but these are distinct notions. Anonymity is preserved when a person's identity in one context is not known in another. Anonymity can be a means to privacy or to avoid harm to oneself e.g. reputational damage. Indeed, anonymity is vital in some situations, for example in the case of an undercover operative whose real identity might be revealed to the criminal organisation he has infiltrated by using facial recognition technology to search billions of facial images on the Internet, social media and elsewhere that were originally created some years earlier when he worked as a uniformed police officer. Such examples demonstrate that anonymity is sometimes an instrumental good. But they do not demonstrate that it is a constitutive human good. In this respect anonymity is quite different from privacy.

The sphere of individual privacy can be widened to include other individuals who stand in a professional relationship to the first individual, for example, a person's doctor. Moreover, morally legitimate institutional processes give rise to confidentiality requirements with respect to information. For instance, law enforcement operations give rise to stringent confidentiality requirements, given what is often at stake, e.g. the outcome of important investigations that could be compromised by exposure or, as mentioned above, the risk to an undercover operative if their identity is revealed [21]. At least in the case of security agencies, such as police, military and intelligence agencies, a degree of compliance with principles of confidentiality is a constitutive institutional good in the sense that security agencies could not successfully operate without a high degree of confidentiality.

Confidentiality is often referred to as informational *security*. So, confidentiality is a species of security. Moreover, confidentiality is, as we saw above, often based on privacy, e.g. the confidentiality of personal information. Accordingly, not only is privacy not necessarily in conflict with security: privacy quite often depends on security. On the other hand, the integration or interlinking of databases of confidential information is potentially problematic from a privacy and autonomy perspective, as the example of the surveillance system in China described above demonstrates.

Another related notion of interest to us here is secrecy [3]. Secret information is not necessarily challenged by the moral right to privacy or by the principle of confidentiality. For unlike privacy and confidentiality, secrecy is a morally neutral or even pejorative notion. Secrecy is at home in contexts of conflict and fierce competition, for example wars, organised criminality and market-based companies. More generally, secrecy is at home in contexts of security. However, high levels of secrecy can mask incompetence, corruption, illegality and human rights abuses, for example in authoritarian regimes. Also, as mentioned above, even in liberal democracies there is

the risk that if the use of the database is not closely monitored and transparent then it will be used for unintended purposes such as surveillance, and, thereby, enable function creep. Accordingly, in contrast with confidentiality, secrecy is not a constitutive institutional good.

We have distinguished privacy, autonomy, anonymity, confidentiality and secrecy, and argued that whereas privacy is a constitutive human good—in part by virtue of its relation to autonomy—and confidentiality a constitutive institutional good, neither anonymity nor secrecy are constitutive goods [20]. Given the close relationships between privacy and confidentiality, on the one hand, and between confidentiality and security, on the other hand, the sharp contrast often drawn between privacy and security does not necessarily obtain.

The notion of security is somewhat vague. Sometimes it is used to refer to a variety of forms of collective security, for example national security (such as harm to the public from a terrorist attack), community security (such as in the face of disruptions to law and order posed by violent political demonstrations) and biosecurity (such as threats to public health and society caused by COVID-19). At other times it is used to refer to personal physical security.

Aside from questions about the scope of security, (for example the personal, organisational and national levels), security can be distinguished by type. Here a distinction between informational and non-informational security can be helpful. Informational (or data) security, as mentioned above, basically consists in ensuring that personal and other confidential information are protected from unauthorised or otherwise illegitimate access. Encryption (of which more in the following section) plays a key role in ensuring data security. Clearly data security is critical in the face of sustained hacking by state and non-state actors that can compromise privacy and confidentiality. Non-informational security pertains to physical or psychological harm to human beings, damage to physical objects, and certain forms of harm to institutional processes or purposes, for example by means of corruption.

Aside from the scope and types of security there are also various contexts of security. These include crime, counter-terrorism, war, cyberwar, trade ‘wars’ and so on. Moreover, the stringency of privacy rights and confidentiality requirements need to be relativized to context. In wartime, for instance, military intelligence gathering is largely unfettered and the privacy rights of citizens curtailed under emergency powers. By contrast, in domestic law enforcement there is, as we saw above, a strong presumption in favour of the privacy rights of citizens. Moreover, in domestic law enforcement there is likely to be increased accountability when privacy rights are overridden. For instance, police might not be able to sign off on access to personal information; rather a judicial warrant might be required. Counter-terrorism in well-ordered jurisdictions is typically a matter of law enforcement. However, in war zones, such as combating Islamic State in Iraq and Syria, counter-terrorism operations, including intelligence gathering, are military in character [19]. Let us now turn to cryptographic technologies, an understanding of which is necessary if we are to offer a coherent account of the ethical problems in this area.

## 2 Encryption

Modern computer-based cryptography comes in a number of methodologies, e.g. public/private key (PPK) cryptography. For our purposes here, we first need to distinguish between passwords and keys. A password can be thought of as an access mechanism; a key is used in an encryption algorithm. Passwords are often quite short, e.g. eight characters. Being short, passwords are susceptible to brute force attack; an attack in which every possible combination is tried in succession, until the solution is found. Thus, protection from unauthorised access is often afforded by a mechanism which wipes all content on the device after, say, 10 attempts to find the password, as in the case of the iPhone of the terrorist Farook mentioned above. By contrast, keys are a lot longer—the longer the better—and are sometimes retrieved by user entered passwords. Accordingly, even in the case of encrypted material there is potentially a weak link in the chain, namely, the password; depending, of course, on the strength of the password and how securely it is held, (e.g. not written down and pasted on one's computer!) Note that a password when it is sent over the net, to say a bank website, is encrypted by the web browser, typically using strong keys. Whereas we would think of a password in terms of the number of characters, the length of a key is usually given in bits. A bit is the information in a binary (two option) choice, a logical yes or no. Thus, a bit can be represented as a zero or one and we could write the key as a series of zeroes or ones. Since a character is normally 8 bits we could think of a 2048 bit key as equivalent to 256 characters (i.e.  $8 \times 256$ ).

It is important to distinguish encryption of documents and data on a device, such as a phone, from encryption in transmission. The first involves some sort of encryption control, of which a password is the most well-known, but there are other options, such as fingerprint, retinal scan, and so on. Despite ongoing efforts on the part of cyber-security personnel to promote the importance of password protection, people persist in using easy-to-guess passwords, which are thus easy to remember, the name of the dog, house address, favourite fruit, etc. A brute force attack on a password (testing every possibility) requires time proportional to  $m^n$  where  $m$  is the number of options for a character and  $n$  is the number of characters. Thus an 8-character password using alphanumeric characters (the integers 0–9 and the 26 letters of the alphabet in both lower and upper case) gives rise to  $62^8$  possibilities i.e. 200 trillion—which a desktop computer could run through in a relatively short time. If we use the most widely used mapping of letters, numbers and symbols to bit patterns, i.e. the whole extended ASCII<sup>2</sup> character set of 256 characters, we get  $256^8$  possibilities, i.e. millions of trillions. So the number of possibilities is a function not only of the length of the password but also of the number of available characters, although, since the number of characters appears in the exponent, increasing the number of characters is usually a more effective way of increasing password strength. However, there

---

<sup>2</sup> ASCII stands for American Standard Code for Information Interchange. Computers can only understand numbers, so an ASCII code is the numerical representation of a character such as 'a' or '@' or an action of some sort. ASCII was developed a long time ago and now the non-printing characters are rarely used for their original purpose.

needs to be very large numbers of possibilities to defeat even a standard desktop computer. On the other hand, there can be *very* large numbers of possibilities which a standard computer would take decades to run through. Brute force attacks, in which every possibility is tested in sequence or at random, on common standards such as AES would take forever. But encryption may be broken on a much smaller timescale through two mechanisms: the advent of new technology; or new algorithms which test possibilities in some special order or apply some novel filtering. Moore's law, the doubling of computing power every two years has held since 1965 for current silicon. Yet an example of a novel technology is quantum computing, which is rapidly developing at the time of writing, where it has been known since 1999 when Peter Shor's now famous 1999 algorithm demonstrated huge potential speedup from quantum computers for prime factorisation and discrete logarithms [29]. An example of new software attacks came in a series of novel attacks on AES-256, summarised by cryptographer Bruce Schneier<sup>3</sup> *This new attack, by Alex Biryukov, Orr Dunkelman, Nathan Keller, Dmitry Khovratovich, and Adi Shamir, is much more devastating. It is a completely practical attack against ten-round AES-256—One of our attacks uses... 2<sup>39</sup> time to recover the complete 256-bit key... where the best previous attack required 2<sup>120</sup> time.*

Estimating the time for an actual computer to crack a key by brute force obviously depends upon the rapidly growing speed of computers. Nevertheless, MIT physicist Seth Lloyd estimated an upper bound to the speed of a 1 kg laptop based on the laws of physics as they stand today [15]. His ultimate laptop would take about a microsecond to break AES 128. It would take an ultimate computer the size of the Earth about a year to crack AES 256. Needless to say, we don't expect to have ultimate computers any time soon.

If we want to send the document over a public channel we need a good password, obviously, and the recipient needs to have learned this password in some way (such as Diffie Hellman, which we discuss below). However, if we use public private key cryptography (PPK), with, say, a typical key of 2048 bits, 2<sup>231</sup> possibilities, then the communication is even stronger. If somebody intercepts the document, this is the strength of encryption with which they have to deal. The password stays on the device and is not transmitted. The alternative to encrypting the document and sending it over a public channel, is to use an encrypted channel, such as WhatsApp. Any useful channel has to be end-to-end encrypted, meaning that is encrypted on the source device and not decrypted until it gets to the destination device. To avoid key compromise by some means, systems such as WhatsApp use ephemeral keys, into more detail of which we go below.

It is important to distinguish between the interception of communications in real time and the accessing of stored material, including documents. Stored material, even if encrypted, is susceptible to accessing if the device is retrieved by investigators and its password determined. Real-time interception of, and access to, the content (as opposed to the metadata, e.g. time, date, location, sender and receiver of call) of communications protected by end-to-end encryption will be extraordinarily difficult

---

<sup>3</sup> [https://www.schneier.com/blog/archives/2009/07/another\\_new\\_aes.html](https://www.schneier.com/blog/archives/2009/07/another_new_aes.html) Accessed.

unless the communication is intercepted prior to encryption or after decryption. This is because the required decryption is extraordinarily difficult, absent access to encryption keys. (For more details on this see below). Crucially, the encryption keys used for communications in devices using end-to-end encryption are typically ephemeral; they are only used for a single message transmission and then discarded. Accordingly, since WhatsApp, for instance, uses end-to-end encryption, security agencies cannot usefully wire-tap phones using WhatsApp, since anything they acquired would not be decryptable.

Typically, encryption keys resist brute force attacks by virtue of the vast number of possibilities that would have to be tried in the time period available, e.g. a number of possibilities of such magnitude that it would take even a high-powered computer decades to find the correct one. Thus, the RSA algorithm used in PPK requires two very large prime numbers,  $p$  and  $q$ , which are multiplied together to produce an even bigger number  $N = pq$ . Take a number such as 1333. This factorises into 31 times 43, which are both prime numbers. The important thing to know is that as the numbers such as 1333 get bigger, it becomes very difficult to find the constituent primes (31 and 43). The idea is to make  $N$  so big, that finding the two prime factors would take an inordinate amount of time. Hence there has been the pressure on governments from law enforcement and security agencies to enforce access to encryption keys.

To allow security agencies to eavesdrop on conversations with WhatsApp and its kin, is rather complicated, owing to the hierarchy of keys of different lifetimes used in the encryption. Thus, let us consider the simpler case of giving security agencies access to private keys, assuming that there are suitable judicial processes to allow access only in case of real need, along the lines already discussed. Storing all these private keys is itself a security risk: they may get leaked, stolen by hackers or just left in unsecured places by defective software due to careless programmers. An alternative is a sort of skeleton private key, sometimes referred to as a backdoor key. The same issue of keeping skeleton key safe applies of course, but there is an additional problem. There is pretty much consensus amongst cryptographers that creating the structure for such backdoor access weakens the encryption, thus making it easier for hackers to break [2, 13, 14].

In the face of this resistance to providing encryption keys to governments, law enforcement's focus has been on finding passwords or on means of attack that do not rely on decryption by virtue of knowing the keys, but rather on bypassing the keys, e.g. by inserting malware into devices as happened in the EncroChat case (described above). There is also, of course, the possibility of legislation, such as exists already in the UK, where a warrant can be obtained to compel a suspect to decrypt a document with prison terms for non-compliance.

Of course, we will not know for some time exactly how EncroChat was compromised, since the security agencies are hardly likely to divulge this information. The consensus seems to be that this was not a defeat of the encryption but the capturing of messages before they were encrypted and sent, through spyware, which had got into the phone. It was most likely downloaded from EncroChat servers, which had got themselves been infected, and then infected phones with something quite ordinary, such as a news release or a software update. One common spyware technique is key

logging. Every key pressed by the user is recorded in some place hidden to the user and sent across the internet to the spyware's owner. Most, if not nearly all, phone apps phone home on a regular basis, usually without the user knowing [32].

The principal encrypted voice call and message systems at the moment are: Signal, Telegram, WhatsApp (owned by Facebook) and Facetime (owned by Apple). Let us consider WhatsApp as illustrative. WhatsApp was very popular, even before it was taken over and became part of Facebook infrastructure. It is end-to-end encrypted, the gold standard, which means that it is encrypted by the sender, decrypted by the receiver and not decrypted anywhere along the way. A highly desirable feature of encrypted messaging is that it should be *open source*. Effectively this means anybody, especially cryptography experts, to scrutinise the details of the algorithms and their implementation. WhatsApp was developed from Signal, using the so-called Signal protocol, and Signal is open source. WhatsApp is not. However, despite recent controversy over the sharing of its *metadata* with parent company Facebook, the best available evidence is that it is still end-to-end encrypted. The EFF (Electronic Frontiers Foundation, one of the leading advocates for technology supporting freedom and justice) states in January 2021 that<sup>4</sup> *To be clear: WhatsApp still uses strong end-to-end encryption, and there is no reason to doubt the security of the contents of your messages on WhatsApp.*

Of course, the provider could have a system in which they keep the encryption keys and save the messages, which means that the message could be decrypted by a third party at a later date. As discussed above, law enforcement has supported this since it would be to their advantage. At any rate, to give users confidence in their communications being forever secret, and as we saw above, the app uses ephemeral keys, which are created for a particular message transmission and then discarded. The user's private keys are never sent anywhere and are not known to the provider.

There are basically two approaches to encrypting a document: block ciphers, such as AES, which break the document up into chunks (blocks) and encrypt each individually; and stream ciphers such as RC4 (Rivest Cipher 4, after its inventor), which operate one character at a time.

Today's block ciphers are both very complicated and very secure. The data is broken up into blocks. Each sub-block is individually encrypted using algorithms then combined with other blocks and the process repeated for a dozen or so iterations. The current more secure version is AES256.

Stream ciphers date back to the sixteenth century with the invention of the one-time pad, beloved of espionage stories ever since. The pad is some document, say Tolstoy's book, *War and Peace*. Starting at some agreed place in the book (our spies have to agree on the book and where to start) the message is compared letter by letter with the book and some reversible algorithm is used to go from one to the other. Thus, if the message has a  $k$  and the book at the same point has a  $q$ , then the algorithm would output, say, a  $z$ . Going backwards taking the  $z$  in the encrypted document, comparing it with the  $q$  in the book spits out  $k$ . The algorithm commonly used is XOR. The computational equivalent is the Vernam cipher which combines the characters of a

---

<sup>4</sup> <https://www.eff.org/deeplinks/2021/01/its-business-usual-whatsapp> Accessed.

document one by one with a random character from the keystream (the letters one by one from the book in our Tolstoy example). The one-time pad and consequently the Vernam Cipher were shown by Claude Shannon to be unbreakable, given that the one-time pad is perfectly random [27]. In the Vernam cipher we use a keystream, which is just a random series of characters. Computer random number generators are now very good at producing very long strings of integers/characters with no relationships between them and no recurring patterns of any kind. But they are only ever pseudo-random. The generator will have control parameters and a starting state, and, if these are replicated, the replica will enable the production of exactly the same sequence. As is obvious, in the pre-digital computing days of cryptography keeping the code book secure was vitally important. Of course, with the advent of keystream (Vernam) ciphers, the code book has been replaced by a random number generator. However, it is now vitally important to keep the details of its parameters and starting state (though not necessarily its algorithm) secure.

An essential point to note here is that cryptographic systems may fail for three reasons: computer power increases allowing a brute force attack (essentially working through every possibility, as mentioned above); the invention of new attack algorithms, or hardware, such as quantum computers; and simply flaws in implementation.

The most effective attacks are not brute force, but exploit some loophole in the cryptography design. Mostly the problems are in software, but occasional a bug appears at the hardware level. This year *The Verge* reported on a particularly nasty vulnerability in Intel chips, which could enable the construction of key loggers, referred to above:

Security firm Positive Technologies discovered the flaw, and is warning that it could break apart a chain of trust for important technology like silicon-based encryption, hardware authentication, and modern DRM protections. This vulnerability jeopardizes everything Intel has done to build the root of trust and lay a solid security foundation on the company's platforms, explains security researcher Mark Ermolov. [35]

Such hardware vulnerabilities are extremely hard to fix (in the worst case requiring chip replacements) [1]:

These types of attacks, called Meltdown and Spectre, were no ordinary bugs. At the time it was discovered, Meltdown could hack *all* Intel x86 microprocessors and IBM Power processors, as well as some ARM-based processors. Spectre and its many variations added Advanced Micro Devices (AMD) processors to that list. In other words, nearly the whole world of computing was vulnerable... ..fixing these vulnerabilities has been no easy job.

Of course, programmers can make errors in implementing cryptographic algorithms. Cryptography is not immune to software bugs.

A fundamental problem in cryptography is agreeing on passwords or encryption keys, using a public channel, where everybody can read the transmissions but cannot infer the password. This is the idea behind a Diffie-Hellman *key exchange* used in PPK and in ECC (elliptical curve cryptography) relied upon by WhatsApp. The following gives a rough idea of how it works.

Xenakis and Zadok want to agree a password. First, they each choose a very large prime number as a private key. Xenakis chooses 43 and Zadok chooses 31.



Now X and Z pick a number, let's say 187. They agree on this over the public channel and again, anybody can know. Now comes the clever trick. X raises 187 to his secret number, 43, getting the very large number.

4888651528060145912868616867727063192303125716802722048864823484528  
9721303752646988922050137964003.

Meanwhile Z does the same with his secret number, 31, getting.

2673559185267605945178503962446826969650755006001031296938716712  
0274163.

X and Z exchange their huge numbers. It doesn't matter if anybody is eavesdropping, since the discrete logarithm problem is hard to solve for them to find either X or Z's secret number. Now each takes the number they receive and exponentiates it with their own secret number. X gets an even bigger number, which would take a page to display. It starts off.

2316655802185836713052880933213078993246302935442089  
4791693836646087967238161954274200463446248956046412  
3889608443987676651933304066297159504611394237176564  
2665535969209484838070647948449175023092257003434334.

Z does the same. She takes the big number she gets from X, call it  $x_1$  and computes  $x_1^{31}$ . Her number begins.

2316655802185836713052880933213078993246302935442089  
4791693836646087967238161954274200463446248956046412  
3889608443987676651933304066297159504611394237176564  
266553596920948483807064794844917502309225700343434

and, in fact, they are *exactly* the same. This huge number is now their shared password. To work out this password from the public traffic, the eavesdropper would need to solve a big discrete logarithm problem.

Let us conclude this section by considering the level of security on Apple devices. Apple has two backup options [36].

1. Via Finder/iTunes, you can turn on encrypted backup (it is off by default). If you do so you need to create a password. But there is no way of using the backup if you lose the password. Thus, you must create a password that you'll remember or you must write it down and store it safely, because there's no way to use your backup without this password.
2. Via iCloud (the default and apple preferred option). Now Apple has the encryption keys. It would argue that this is good for users since if they lose the password, Apple can recover it.

However, although Chinese iPhones will retain the security features that can make it all but impossible for anyone, even Apple, to get access to the phone itself, that will not apply to the iCloud accounts [23]. Any information in the iCloud account could be accessible to Chinese authorities who can present Apple with a legal order. Elsewhere the keys are stored by Apple in the US, which means, under a suitable court order in the US courts, Apple could be forced to give up the keys and hence the data on the phone. Now it seems that WhatsApp messages are backed up to the



cloud unencrypted. From their FAQ, WhatsApp chat histories aren't stored on their servers. Media and messages you back up aren't protected by WhatsApp end-to-end encryption while in iCloud. If you've previously backed up your iPhone using iCloud or iTunes, you might be able to retrieve your WhatsApp chats by restoring your iPhone from a previous backup.

In a strange twist, Google, which depends heavily on targeted advertising revenue, and obtains this through massive surveillance of how its users employ its services, nevertheless offers greater personal security than Apple. Data backed up to Google is encrypted by a key, accessed by the phone's pin number or fingerprint etc., and this key is controlled on Google's servers by a custom chip referred to as Titan. Now, since a pin number is a very weak password, the Titan uses the old maximum number of tries principle (although we do not know how many tries this actually amounts to) [10]. The limited number of incorrect attempts is strictly enforced by a custom Titan firmware that cannot be updated without erasing the contents of the chip. By design, this means that no one (including Google) can access a user's backed-up application data without specifically knowing their passcode.

### 3 Ethical Analysis

In the light of our conceptual analysis of privacy, confidentiality, autonomy and security, and our descriptive technical account of encryption, we can now offer an ethical analysis of privacy rights and security needs in relation to encryption in the overall context of the counter-terrorism policies of liberal democratic states. Before addressing the specific issues of privacy and encryption in counter-terrorism, a number of general points that bear on this issue and which are extractable from the discussions in Sects. 1 and 2 need to be made.

We have argued that privacy rights, including in respect of smartphone content and metadata, are important, in part because of their close relation to autonomy. However, we also noted that privacy rights are not absolute; they can justifiably be overridden, for instance, in relation to an imminent terrorist attack. Therefore, the strong claim that some privacy advocates are inclined to make, namely, that there is, in effect, an absolute moral right to very strong, i.e. uncrackable, encryption, since it asserts there are no circumstances in which very strong encryption should be impermissible, is not sustainable. This is, of course, not to demonstrate that very strong encryption is morally impermissible under all circumstances. Perhaps, for instance, citizens who live in an authoritarian state are morally justified in possessing devices equipped with very strong encryption. Moreover, even in liberal democracies very strong encryption might be morally permissible if there were other means by which law enforcement agencies could efficiently and effectively investigate and, if justified, charge terror suspects. For instance, if bulk metadata (as opposed to communicative content) in the context of machine learning techniques combined with other methods, such as hacking and insertion of malware was sufficient (as presumably occurred in the EncroChat scenario). On the other hand, bulk metadata collection and, relatedly, integrated databases, are themselves problematic from a privacy perspective.

Although privacy rights can be overridden under some circumstances, notably by law enforcement investigations of serious crimes including terrorism, there is obviously a point where infringements of privacy rights are excessive and unwarranted. Security agencies' ongoing, ready access to the personal data of the entire population would be clearly unacceptable. Moreover, regulation, and associated accountability mechanisms need to be in place to ensure that, for instance, personal information obtained for a legitimate purpose, such as counter-terrorism, can be accessed by law enforcement officers to enable them to detect suspects and protect citizens from being murdered, but not used to identify protesters at a political rally [25].

We have also argued that the sharp contrast between privacy and security cannot be maintained, since security includes informational or data security, i.e. security of personal data and confidentiality in relation to data held by security agencies. Moreover, it is primarily goods that are not essentially informational that ultimately need to be weighed so as to achieve an acceptable moral equilibrium, notably individual autonomy, personal security and institutional integrity.

Moreover, by describing the technical issues in some detail we have shown how it is that high level end-to-end encryption is, in effect, invulnerable to decryption. However, as we have also shown by describing the technical issues in some detail, how devices that use such encryption are, nevertheless, vulnerable by virtue of their use of passwords and the possibility of being hacked and the insertion of malware.

In the light of the above, a number of interconnected ethical issues have come into view. Some of these arise from the expanding use of bulk data collection and surveillance in counter-terrorism operations, especially in the context of interlinkage of databases, data analytics and artificial intelligence. As already mentioned, these developments are relevant to debates surrounding encryption in so far as they provide an advantage to security agencies that might to some extent mitigate the problem of not having access to encrypted communications and documents.

This is not to say that there ought not to be constraints on bulk data collection and analysis. For instance, it is unacceptable for data, including surveillance data, originally and justifiably gathered for one purpose, e.g. taxation or combating a pandemic, to be interlinked with data gathered for another purpose, e.g. counter-terrorism, without appropriate justification. The way metadata use has expanded from initially being used by only a few agencies engaged in counter-terrorism to now being used quite widely by governments in many western countries, is an example of function creep.

Another important development that needs to be kept in mind when adjudicating privacy and encryption issues in counter-terrorism contexts is the blurring of the distinction between the application of the domestic law enforcement and the military combat frameworks in counter-terrorism operations, given that terrorist organisations, such as Al Qaeda and Islamic State operate in war zones as well as in well-ordered jurisdictions. What are the privacy rights of, for instance, those suspected of travelling abroad with the *intention* of becoming foreign terrorist fighters but who are yet to fulfil this intention? Should they be treated as ordinary citizens possessed of

the full array of privacy and other rights who are only potential, and not actual, criminals?<sup>5</sup> Again, what are the privacy rights of those suspected of being foreign terrorist fighters who have returned to their home country? Should they be treated as ordinary citizens possessed of the full array of privacy and other rights albeit, if returnees, citizens suspected of criminality? Or should they be regarded, in effect, as suspected terrorist-combatants and, therefore, suffer a curtailment of their privacy and other rights even in the absence of sufficient evidence to convict them of terrorist offences, e.g. in relation to privacy, the ongoing monitoring of their private communications by domestic security agencies, the retention of their personal data by domestic security agencies, and the disclosure of this data to third parties such as foreign governments and their security agencies [33].

Finally, it should be noted that there is a danger in relation to the technological developments discussed here (e.g. bypassing encryption and the use of integrated bulk databases), as there is in relation to technological developments discussed elsewhere (e.g. the use of facial recognition technology) [30], that various general principles hitherto taken to be constitutive of liberal democracy are gradually undermined, such as the principle that an individual has a right to freedom from criminal investigation or unreasonable monitoring (including accessing of the content of their communications), absent prior evidence of violation by that individual of its laws. In a liberal democratic state, it is generally accepted that the state has no right to seek evidence of wrongdoing on the part of a particular citizen or to engage in selective monitoring of that citizen, if the actions of the citizen in question have not otherwise reasonably raised suspicion of unlawful behaviour and if the citizen has not had a pattern of unlawful past behaviour that justify monitoring. However, this principle is potentially undermined by certain kinds of offender profiling and, specifically, ones in which there is no specific (actual or reasonably suspected) past, imminent or planned crime being investigated. We note that not simply communicative content but also meta-data could be used for profiling, risk assessment and monitoring of people who are considered at risk of committing crimes. Moreover, in a liberal democratic state, and related to the above-mentioned principle, there is a general presumption against the state monitoring the citizenry. This presumption can be overridden for specific purposes but only if the monitoring in question is not disproportionate, is necessary or otherwise adequately justified and kept to a minimum, and is subject to appropriate accountability mechanisms.

In this chapter we have performed three main interconnected tasks. First, we have offered an analysis of the nature and moral significance of privacy, including its relationship to confidentiality, autonomy and security, in the context of the counter-terrorism responses of liberal democratic states. Second, we have provided a description of relevant cryptographic technologies. One focus here has been on WhatsApp. Third, we have discussed the privacy rights and security needs in relation to encryption in the overall context of the counter-terrorism policies of liberal democratic states.

---

<sup>5</sup> Although in some jurisdictions, such as Australia, travelling to Syria and other zones of armed conflict is in and of itself a crime. See Section 119.2 of the Criminal Code of Australia.

**Acknowledgements** I, Seumas Miller (co-editor), would like to thank for the funding from the ERC Advanced Project on Collective Responsibility and Counter-terrorism (of which I am the Principal Investigator).

## References

1. Abu-Ghazaleh N, Ponomarev D, Evtushkin D (2019) How the spectre and meltdown hacks really worked. *IEEE Spectrum*. <https://spectrum.ieee.org/computing/hardware/how-the-spectre-and-meltdown-hacks-really-worked>. Accessed 24 Sept 2020
2. Benaloh J (2018) What if responsible encryption back-doors were possible? *Lawfare*. <https://www.lawfareblog.com/what-if-responsible-encryption-back-doors-were-possible>. Accessed 29 Sept 2020
3. Bok S (1982) *Secrets: on the ethics of concealment and revelation*. Pantheon Books, New York
4. Botelho G, Ellis R (2015) San Bernardino shooting investigated as ‘act of terrorism’. *CNN*. <https://edition.cnn.com/2015/12/04/us/san-bernardino-shooting/index.html>. Accessed 5 May 2021
5. Cox J (2020) How police secretly took over a global phone network for organized crime. *Vice Motherboard*. <https://www.vice.com/en/article/3aza95/how-police-took-over-encrochat-hacked>. Accessed 2 July 2020
6. Fried C (1969) Privacy. *Yale Law J* 77(3):475–493
7. Grossman L (2016) Inside Apple CEO Tim Cook’s fight with the FBI. *Time*. <https://time.com/4262480/tim-cook-apple-fbi-2/>. Accessed 5 May 2021
8. Henschke A (2017) *Ethics in an age of surveillance: virtual identities and personal information*. Cambridge University Press, New York
9. Inness JC (1992) *Privacy, intimacy, and isolation*. Oxford University Press, New York
10. Jonnalagadda H (2020) Apple may have ditched encrypted backups, but Google hasn’t. *Android Central*. <https://www.androidcentral.com/apple-may-have-ditched-encrypted-backups-google-hasnt>. Accessed 28 Sept 2020
11. Kleinig J, Mameli P, Miller S, Salane D, Schwartz A (2011) *Security and privacy: global standards for ethical identity management in contemporary liberal democratic states*. ANU Press, Canberra
12. Koops B-J, Newell BC, Timan T, Škorvánek I, Chokrevski T, Galič M (2016) A typology of privacy. *Univ Pennsylvania J Int Law* 38(2):483–575
13. Landau S (2018) Exceptional access: the devil is in the details. *Lawfare*. <https://www.lawfareblog.com/exceptional-access-devil-details-0>. Accessed 29 Sept 2020
14. Landau S (2020) If we build it (they will break in). *Lawfare*. <https://www.lawfareblog.com/if-we-build-it-they-will-break>. Accessed 29 Sept 2020
15. Lloyd S (2000) Ultimate physical limits to computation. *Nature* 406(6799):1047–54
16. Lucas GR (2013) Privacy, anonymity, and cyber security. *Amsterdam Law Forum* 5(2):107–114. <https://doi.org/10.37974/ALF.253>
17. Macnish K (2018) Government surveillance and why defining privacy matters in a post-Snowden world. *J Appl Philos* 35(2):417–432. <https://doi.org/10.1111/japp.12219>
18. Matthews S (2010) Anonymity and the social self. *Am Philos Q* 47(4):351–363
19. Miller S, Feltes J, Henschke A (2021) *Counter-terrorism: the ethical issues*. Edward Elgar, London
20. Miller S, Walsh P (2016) NSA, snowden and the ethics and accountability of intelligence gathering. In: Galliot J, Reed W (eds) *Ethics and the future of spying: technology, intelligence collection and national security*. Routledge, London, pp 193–204
21. Miller S, Gordon I (2014) *Investigative ethics: ethics for police detectives and criminal investigators*. Wiley-Blackwell

22. Nakashima E (2016) Apple vows to resist FBI demand to crack iPhone linked to San Bernardino attacks. The Washington Post. [https://www.washingtonpost.com/world/national-security/us-wants-apple-to-help-unlock-iphone-used-by-san-bernardino-shooter/2016/02/16/69b903ee-d4d9-11e5-9823-02b905009f99\\_story.html](https://www.washingtonpost.com/world/national-security/us-wants-apple-to-help-unlock-iphone-used-by-san-bernardino-shooter/2016/02/16/69b903ee-d4d9-11e5-9823-02b905009f99_story.html). Accessed 5 May 2021
23. Nellis S, Cadell C (2018) Apple moves to store iCloud keys in China, raising human rights fears. Reuters. <https://www.reuters.com/article/us-china-apple-icloud-insight/apple-moves-to-store-cloud-keys-in-china-raising-human-rights-fears-idUSKCNIG8060>. Accessed 28 Sept 2020
24. Nissenbaum H (2009) Privacy in context: technology, policy, and the integrity of social life. Stanford Law Books
25. Robbins S (2021) Bulk metadata collection and the right to privacy. In: Miller S, Feltes J, Henscke A (eds) Counter-terrorism: the ethical issues. Edward Elgar, London
26. Schmidt MS, Pérez-Peña R (2015) F.B.I. Treating San Bernardino attack as terrorism case. New York Times. <https://www.nytimes.com/2015/12/05/us/tashfeen-malik-islamic-state.html>. Accessed 5 May 2021
27. Shannon CE (1945) A mathematical theory of cryptography. Index PO. 4, Memorandum for file, case 20878, MM 45-110-92
28. Shaw D (2020) Hundreds arrested as crime chat network cracked. BBC News. <https://www.bbc.com/news/uk-53263310>. Accessed 2 July 2020
29. Shor PW (1999) Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev 41(2):303–332
30. Smith M, Miller S (2021) The ethical application of biometric facial recognition technology. AI Soc. <https://doi.org/10.1007/s00146-021-01199-9>
31. Solove D (2008) Understanding privacy. Harvard University Press, Harvard
32. Swinhoe D (2018) What is a keylogger? How attackers can monitor everything you type. CSO. <https://www.csoonline.com/article/3326304/what-is-a-keylogger-how-attackers-can-monitor-everything-you-type.html>. Accessed: 29 Sept 2020
33. UNSC (2015) Gaps in the use of advance passenger information and recommendations for expanding its use to stem the flow of foreign terrorist fighters, 26 May 2015, S/2015/377, para 44
34. Warren SD, Brandeis LD (1890) The right to privacy. Harv Law Rev 4(5):193–220
35. Warren T (2020) A major new Intel processor flaw could defeat encryption and DRM protections. The Verge. <https://www.theverge.com/2020/3/6/21167782/intel-processor-flaw-root-of-trust-csmesecurity-vulnerability>. Accessed 24 Sept 2020
36. WhatsApp LLC (2020) How to back up to iCloud. WhatsApp Web. <https://faq.whatsapp.com/iphone/chats/how-to-back-up-to-icloud/>. Accessed 28 Sept 2020

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



# An End to Encryption? Surveillance and Proportionality in the Crypto-Wars



Kevin Macnish

**Abstract** End-to-end encryption has been a reality for at least 30 years. However, it is only with recent developments that it has become widespread on mobile phones operating over the internet. This has provided tools for terrorists to plan activities that lead directly to the deaths of innocent civilians. At the same time, it has also been used by dissidents challenging totalitarian regimes and holding liberal democracies to account. In this chapter I argue that while terrorist use of such encryption may render that encryption unjustifiable within a liberal democracy, within an international context the protection that it provides to those seeking to establish law-abiding democracies is too great to be ignored.

## 1 Introduction

The encryption of messages, and subsequent attempts to decrypt the same by people other than the intended recipient, is an ancient practice [49]. Throughout the twentieth century, this became increasingly digitized as the Cold War and technological developments led to countries investing in increasingly complex methods of encryption, both in terms of hardware and software [5, 15]. With the introduction of the internet and the widespread use of smart mobile telephony, though, encryption has entered a new phase [4]. In the past, encryption by anyone other than a state actor has largely been a matter of amateur interest and ability, with most people lacking the mathematical skills and computing power to engage in complex cryptography. Recent developments, though, now mean that anyone in possession of a smart phone can encrypt their communications to a level of complexity matching that of many states' own cryptographic agencies [19, 45].

These developments come at a price. What may be seen as the democratization of encryption can also be viewed as the protection of immoral activity, not least

---

K. Macnish (✉)

Sopra Steria and University of Leeds, 1 Bartholomew Close, EC1A 7BL London, England

© The Author(s) 2021

A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,  
Advanced Sciences and Technologies for Security Applications,  
[https://doi.org/10.1007/978-3-030-90221-6\\_10](https://doi.org/10.1007/978-3-030-90221-6_10)

155

terrorism. To some degree, this protection has been limited by the technology available to the average person. For the most part, this has been “transport layer” encryption, such as Secure Socket Layer encryption (SSL—recognized online through URLs beginning “https://” vice “http://”). This functions by encrypting a message between sender and hosting platform, typically a website, which then decrypts the message. If the message is forwarded to a recipient other than the hosting platform, that host then re-encrypts the message and sends it on. Hence transport layer encryption allows a degree of encryption, but a hosting platform can always access the plain (decrypted) text of the message.

For example, if I were to send an unpublished, sensitive research paper to a colleague, I might use a web-based email service. These services are usually encrypted with SSL. By using this service, the paper would be encrypted on leaving my computer so that anyone intercepting my communications would be unable to read it. On reaching the email service’s web server, though, it would be decrypted and stored. When my colleague then opened her email to read the paper, it would then be re-encrypted by the server and sent to her computer. Thus, the paper is encrypted while in transit across the web, but not when it is at rest on my computer, my colleagues’ computer, or, perhaps most significantly, on the email service’s web server.

End-to-end encryption (E2EE), in contrast, cannot be read by anyone other than sender and intended recipient. The platform over which it is sent cannot access the content, thus guaranteeing an additional degree of protection for the secrecy of the message. E2EE frequently uses Public Key Encryption techniques in which sender and recipient each have a public and a private key through which they can encrypt and decrypt each other’s messages (e.g. Pretty Good Privacy—PGP).

Returning to the above example of sending a research paper over a web-based email service, I may decide that I am unhappy with the potential of that service being able to access and read my paper. Perhaps the paper is critical of that service, or it contains commercially sensitive information. In that case, I would encrypt it while it is still on my computer using a combination of a session key and my colleague’s public key, which is unique to her but publicly available. I could then send the encrypted paper to my colleague for her to decrypt using her private key, which, as the name suggests, is unique to her and not publicly available [32, 104]. In this way, the research paper remains encrypted while in transit across the web and while resting on the email service’s web server.

PGP is widely attributed to Phil Zimmermann, who developed it in 1991 [21, 32]. However, James Ellis, an employee at the UK signals agency GCHQ, developed the concept behind PGP in the 1970s but, due to the potential for abuse and the possibility of hostile actors having encryption which GCHQ was unable to decrypt, decided not to share it publicly [22, 23]. Even after Zimmermann popularised the concept, though, take-up of PGP was limited to those prepared to take the time and effort to understand and implementing the approach. More recently, though, E2EE has been embraced by a number of applications, including Signal, Wickr, WhatsApp, Facebook Messenger, Zoom, iMessage, and Telegram. This has enabled E2EE to



become far more widely implemented through a process largely invisible to the user, and thus requiring no more effort than using a non-encrypted app on a mobile phone.

These concerns render E2EE an issue for national security and law enforcement, as ease of use entails the likelihood that such techniques will be employed by terrorists and other criminal actors. E2EE would render their communication secure from government interference, limiting the state's capacity to identify and prevent attacks. Governments have responded by attempting to curtail the functionality of E2EE accordingly. In the mid-1990s, the Clinton administration responded to the introduction of PGP with a proposal for the "Clipper Chip", which would provide a back door into encrypted communications for the state. That is, however good the encryption, the state would be able to access a master key which would break any encrypted media. Access to this "backdoor" into encrypted communications would be controlled by a trusted third party, holding the key in escrow. However, a significant public backlash in what became known as the crypto-wars led to the proposal being abandoned in 1996 [44, 52, 68].

In this chapter, I examine the evidence for terrorist use of E2EE and the counter-measures proposed by government. I then consider the debate in the standard terms of privacy versus security. This, though, leads to an undervaluing of the core interests that privacy protects, particularly security. After a brief discussion on the different aspects of security, I argue that there is a proportionality consideration that needs to be made in the E2EE debate. In conducting this, I conclude that while E2EE is arguably of less need in constitutionally robust and accountable liberal democracies, the international nature of digital communications mean that the loss of E2EE would endanger movements for democracy and the rule of law around the globe. This, I hold, is a price not worth paying.

## 2 Terrorist Use of E2EE

The concern regarding the threat posed by E2EE has traditionally been presented by the state in terms of terrorism [81, 85], and there is certainly evidence of terrorists using the technology.<sup>1</sup> In 2007, al Qaida created a custom-built encryption tool called "Mujahedeen Secrets", which was known to be used in 2009 to communicate with Western-based operatives [89] and the use of which was taught to German recruits planning an attack in Europe the following year [14]. Al Qaida in the Arabian Peninsula cleric Anwar Al Awlaki devised his own custom encryption technique, which he used to communicate with Rajib Karim, a British Airways call centre worker in plotting the destruction of an aircraft and discussing the recruitment of further sympathisers [18].

With Islamic State (IS) the use of encryption turned from custom-made solutions to commercial off the shelf (COTS) products. While there is debate as to the degree

---

<sup>1</sup> Note that I am here taking "terrorism" at face value to be an indiscriminate threat to civilian life, rather than a cynical labelling in international politics used to distinguish friends from enemies.



to which encryption tools such as WhatsApp were instrumental in the plotting of major terrorist incidents, such as the Paris attacks in 2015, IS sympathisers have been reported as using a raft of COTS products, including Kik, PGP, Surespot, Tails, Telegram, Tor, TruCrypt, WhatsApp, and Zello, for their communications, including sharing beheading videos, developing networks, and proffering advice on how to hide one's presence on the internet [29, 98]. Surespot was favoured by British IS operative Junaid Hussain to share bomb-making tips, and he is known to have used the tool to discuss targeting options with Junead Khan, another British extremist who was convicted of a plot to attack US military personnel in the UK [13, 17, 102]. Hussain also used an unspecified encryption system to communicate with a US sympathiser of IS who was in a group that shot at a "Draw the Prophet Mohammed" contest in Texas in May 2015, exchanging 109 encrypted messages on the day of the attack [38, 84].

It is therefore reasonable to conclude that terrorists are indeed using E2EE technologies, which provide "safe spaces" for them to communicate online. There is as yet no concrete public evidence of their having used these to plan major events, although this would be difficult to establish given the nature of the encryption. Furthermore, it is noteworthy that the bulk of the publicly known use of E2EE by terrorist operatives, as opposed to their sympathisers, can be isolated to two users: Al Awlaki and Hussain. However, there is sufficient evidence to say that at least some terrorists have used E2EE to network and recruit people to their cause, and it is very likely that some have also discussed forthcoming attacks using the cloak of E2EE.

Assuming a broadly Lockean notion of a central responsibility of the state being to safeguard its citizens from attack, the concern of governments with terrorist use of E2EE is understandable. Essentially, governments are seen to fail in their primary responsibility every time a terrorist attack takes place in their territory. At the same time, we do not generally hold that the state has *carte blanche* to protect its citizens come what may. In liberal democracies at least, the state is rightly limited in its actions by human and civil rights which protect citizens from over-reach by the state. Furthermore, many if not most liberal democratic governments genuinely wish to protect their citizens' rights and guard against overreach, for fear, if nothing else, of where this could lead with subsequent, less well-meaning governments. Hence, we do not generally see all-out assaults on every conceivable civic right by the state. However, E2EE is an effective way of guaranteeing the privacy of communications, and so responses to E2EE which involve governments reading, or being able to read, encrypted messages directly threaten the right to privacy. To this end, liberal democratic states have sought means of overcoming the limitations imposed by E2EE, such as through introducing a backdoor, while guaranteeing the privacy of citizens' communications.

### 3 Countering E2EE

Arguments by the state for the insertion of backdoors such as the Clipper Chip into E2EE are clear, and have been made into law in some countries such as Australia [27]. However, once it is known that an encryption system has a backdoor this will render the system a likely target for attack by others, both malicious and security-minded actors seeking to find the backdoor and thereafter exploit or demonstrate it as a weakness in the system. Essentially, once there is a backdoor, it becomes very difficult to stop the “wrong” people gaining access. For this reason, Alex Stamos, former chief security officer at Yahoo, has likened backdoors to “drilling a hole in the windshield” such that the backdoor fatally undermines the integrity of the whole system [71]. This argument was used in 2015 when, following the San Bernadino terrorist shootings when the FBI took Apple to court in an attempt to force the technology company to provide a backdoor to its iPhones. Apple’s response was precisely that backdoors would be accessed and abused by malicious actors other than the state, and that the state itself was not entirely to be trusted. CEO Tim Cook argued that, “if a court can ask us to write this piece of software, think about what else they could ask us to write—maybe it’s an operating system for surveillance, maybe the ability for the law enforcement to turn on the camera ... I don’t know where this stops. But I do know that this is not what should be happening in this country” [28].

More recent attempts have been made to undermine the power of E2EE, notably but not exclusively in the US. These include the Eliminating Abusive and Rampant Neglect of Interactive Technologies (EARN IT) bill, introduced in March 2020 to the U.S. Senate by Republican Senator Lindsey Graham and Democratic Senator Richard Blumenthal, and the Lawful Access to Encrypted Data (LAED) bill, introduced in June 2020 by Republican Senators Lindsey Graham, Tom Cotton and Marsha Blackburn. These are the latest incarnations of what has been dubbed Crypto-war 2.0, in reference to the earlier debate surrounding the introduction of PGP and the proposed Clipper Chip [44, 52, 68].

If EARN-IT is passed, companies would become liable for terrorist communications taking place over their platforms if it can be demonstrated that they failed to take adequate measures to protect against this. Allowing or encouraging E2EE could then be seen as precisely this sort of failure [3]. The LAED bill, if enacted, would “would authorize courts to issue search warrants that would compel ‘a device manufacturer, an operating system provider, a provider of remote computing service, or another person to furnish all information, facilities, and assistance necessary to access information stored on an electronic device or to access remotely stored electronic information, as authorized by the search warrant’” [67]. Hence Facebook, which owns WhatsApp, could be served a warrant to give unencrypted communications placed over WhatsApp to law enforcement. This would require Facebook to hold a backdoor to WhatsApp’s E2EE and so amounts to an effective repeat of the Clipper Chip debate.

As noted, these US efforts are not isolated cases. The LAED bill, if passed, would become the American equivalent of legislation which has existed in the UK since the

passing of the Investigatory Powers Act in 2016 and in Australia since the passing of the Assistance and Access Act in 2018. However, US law carries somewhat greater force in this arena, given that so many technology companies are based in the US and come under US jurisdiction. By contrast, if these companies simply offer services in the UK and Australia but remain headquartered elsewhere, it becomes harder to enforce legislation demanding that the company breaks its own encryption.

If these bills are not passed in the US then E2EE remains, somewhat cynically, in the companies' interests. This is because it allows them to avoid having to deal with government warrants and court orders which are passed to enable the state to forcibly access their content. It would enable technology companies to argue that, unlike SSL, with E2EE they are completely unable to access the content and so unable to comply with any such legal requirement for technical reasons. Furthermore, this would play well to customers who do not trust the state and seek encryption for legitimate communications which they wish to remain secret. To this end, it is notable that the Snowden revelations were a catalyst in encouraging the development of E2EE for public use, adding fuel to the fire for those who distrust state intentions in this area [101].

Even if the introduction of these laws is successful, though, as with the UK and Australia their impact will necessarily be limited in an international marketplace and on the internet. While the US may be able to regulate Facebook, for instance, many of the E2EE services known to be used by terrorists are free software that is openly available online. If any limitations were made to one service under a particular jurisdiction, the likely response by terrorists would be to switch to an alternative platform under a different (or no) jurisdiction [39]. This was raised in 1993, when cryptographer Whitfield Diffie testified to Congress that backdoors would weaken the value of US encryption providers in the global market. The known existence of backdoors in US encryption systems would raise suspicions among potential clients that the US government had access to their communications. Diffie added that those who wanted to hide their communications (i.e. terrorists) could still do so easily, even with such backdoors in place (such as through using code words to mask their activities). The result would be that the only people who would remain susceptible to state surveillance would be those who were not worried about that surveillance. Furthermore, once a backdoor is put in place, Diffie noted, there is no guarantee that others would not be able to exploit the backdoor for their own purposes [31]. Essentially, the same argument as we have seen Tim Cook was to give 23 year later.

Hence technical measures to provide backdoors into E2EE while guaranteeing citizen privacy are at least flawed and arguably impossible without fundamentally undermining the value of E2EE. It is not feasible to maintain E2EE while at the same time enabling governments to break the encryption in their efforts to counter

terrorism. The technology is such that the ability for governments to break the encryption would risk legitimate users having their communications at risk of being intercepted by the state, while terrorists would evade discovery by resorting to alternative means of communication.<sup>2</sup>

The public backlash to the Clipper Chip, Cook's response to the FBI court case, and the catalysing effect of the Snowden revelations all point to the key problem for attempts to counter E2EE, namely that encryption provides a means for legitimate, non-malicious users to encrypt their communications for entirely moral reasons. To this end, appeal is frequently made to the right to privacy in the face of Government surveillance of communications [64], and so it is to an examination of privacy that I turn next.

## 4 Privacy and E2EE

Privacy is widely recognized as a core human right [25, 37, 83, 96], but it is a *pro tanto* rather than an absolute right. As described in human rights declarations and legislation, privacy may be overridden by competing considerations, such as national security and the public interest [62, 64]. Privacy is frequently recognized as a basic need, extending across time and cultures [58] which has both inherent and instrumental value [20, 63] in terms of protecting autonomy [7], governing relationships [78], freedom from embarrassment and freedom to be creative [35]. While each of these is often interpreted in terms of individual benefit, there are significant public benefits from privacy, and not only in terms of the aggregation of individual goods. Privacy is also central to the functioning of the democratic process and society at large [79, 80, 87, 91], and is therefore a key component of liberal democracies. As such, it should not be surrendered lightly.

At the same time, privacy is not a universal good. The claim to privacy can, most obviously, be used to cover illegal and immoral activities, such as planning acts of terrorism [74]. Given the uncontested evils of terrorism, it is hard to argue for the *prima facie* upholding of privacy in this case.<sup>3</sup>

This is particularly true in the case of E2EE, given that a lack of E2EE (or a capability for the government to decrypt E2EE communications) does not automatically entail government mass surveillance, but rather allows for the *possibility* of government surveillance. Hence privacy is put at increased risk by banning or decrypting E2EE but not necessarily lost. The mere *ability* of the state to intercept and "read" the

---

<sup>2</sup> It is also worth pointing out that while it is a significant loss for security services to be unable to access the content of terrorist communications, they retain the in principle ability to perform network analysis on the metadata of these communications. Furthermore, traditional methods of deterrence and investigation continue, and continue to be effective. It would be short-sited to see the breaking of E2EE as the only means to defeat terrorism.

<sup>3</sup> Beyond that, privacy has been criticised by some feminist scholars as demarcating the domestic arena as one in which the state should not intrude, historically leaving men to abuse women and children with relative impunity [1, 34, 54].

communications of its citizens does not automatically entail the *actual* interception and reading of these communications. This may seem like a fine distinction, but it is one which arguably means that the interception of communications which are then not read, or not read by a human, does not entail a violation of privacy [62, 64]. At the same time, there is ample evidence of liberal democratic states having done precisely this in the cases of Martin Luther King, Jr. [33, 66], the Democratic Party leading up to the 1972 US presidential election [24, 36], members of the UK Cabinet in the late 1970s [73], environmental groups in the UK in the 1980s and '90s [57], and Muslim groups in the US post-2001 [99]. None of these groups presented an obvious threat to national security, and yet their communications were accessed nonetheless.

Viewing the E2EE debate in terms of privacy versus security is therefore fraught with difficulty. While providing security is seen as one of the core duties of the state, privacy is a *pro tanto* right which can be overridden in the interests of national security. Furthermore, even if the state is *able* to ban or decrypt E2EE, it does not necessarily follow that privacy *will* be lost. The capacity to intercept and read people's conversations does not necessarily lead to the actual interception and reading of those conversations. I have argued that an interest in privacy extends beyond the individual to the community and ultimately the state, but it is not obvious in and of itself that this interest should extend to allow acts of terrorism to be perpetrated.

To conclude that, in this instance at least, national security interests in countering terrorism should trump those of privacy would be too hasty, though. What is often missed in the privacy versus security debate is that another justification for privacy is that it provides security [60, 63]. While I have privacy, I have security from your (or the state's) intrusion into my life, which is writ large when communities have privacy. This freedom from intrusion provides me/the community with significant security from interference by the state, as seen in the appeal made by numerous US Supreme Court judgments regarding the ability of the state to govern on activities normally engaged in the bedroom [90]. Furthermore, there is a question of distribution that is raised in the privacy vs security debate which can be missed. Precisely whose privacy is being infringed to protect exactly whom? Profiling of terrorist suspects on the basis of ethnicity or religion can often mean that the privacy of minority groups is imperilled while that of the majority remains untouched. This is particularly noteworthy in cases such as the police surveillance of Muslim groups [55, 56, 99] and environmental activists [57].

## 5 Security Versus Security

Rather than viewing the dilemma as one of privacy versus security, then, it may be more constructive to view it instead as one of security versus security. Within the context of counterterrorism, the term "security" is often taken as shorthand for "national security", which gives it a strong force in arguments, such as when pitched against privacy. This is especially so in times of national crisis, such as when a state is facing a particular terrorist threat. However, national security is just one

form of security. Waldron suggests three broad areas of collective security, national security and human security, while Herington identifies a range of uses, including national security, human security, ontological security, emancipatory security and securitization theory [42, 97, p. 459], to which could also be added social security, environmental security, maritime security and doubtless more. Despite this, there are commonalities in all these uses. National security is not so different from social or personal security in that there is a common denominator. Both entail an understanding of exposure to risk such that the greater the exposure, the less the security, and vice versa [65]. My proposal here is therefore to reframe the E2EE debate as concerning the security of the state against the security of the citizens of the state. In both cases, I take security to refer to the preservation of life (human security) and the preservation of a recognized way of life.

E2EE provides protection for citizen security, which may be threatened by individual agents/groups and/or by the state. The first of these, individual agents presenting a threat, may be serious hackers (possibly seeking to access personal details to engage in identity fraud) or amateur actors, such as abusive relatives or stalkers who want to spy on a person. Encrypted communications can help protect citizens from attacks from such actors. E2EE can also protect citizens qua corporations from a risk of cyberattack. By contrast, individuals and corporations *as such* are rarely threatened by acts of terrorism. This is not to say that individuals and corporations are not threatened by acts of terrorism—they are. However, terrorism is to a large extent indiscriminate. Beyond caring that the victims are the “right sort” of victim (i.e. citizens of a particular state), acts of terrorism are rarely directed at particular individuals. As such, E2EE can provide security to citizens against a directed, personalised threat while the loss of E2EE as a means of counterterrorism provides security to citizens against an undirected, impersonal threat.

The second threat to citizen security is from the state. In saying this, it should be remembered that there are more states than just liberal democracies. Totalitarian states clearly threaten the lives and wellbeing of their citizens, should those citizens dissent from the state’s activities. Given the international nature of the internet and encryption, the international audience *must* be considered in the equation. Furthermore, while liberal democracies generally do not threaten the lives and wellbeing of their citizens in the same way as totalitarian states, one of the reasons for this is inherent to the notion of liberalism: that citizens have rights against, and freedoms from, the state. Twentieth century history demonstrates how quickly the state can cease to be liberal or democratic when these rights are removed from citizens, as with the rise of Nazism in Germany and Soviet domination in mid-century Central Europe. It is therefore in the state’s interest to uphold these rights whenever possible.

The rights I have in mind here are those of free speech, free expression, free association, and the right to self-defence (human security). Each of these is recognised in the same international human rights legislation that recognize the human right of privacy. However, unlike privacy, these, and especially the right to self-defence cannot (at least as presented) be overridden by the public interest or national security. Furthermore, each of these rights is strengthened by E2EE. Without reliable encryption, any communications may be discovered and read by a government. While this

may not be a concern for most domestic communication, even liberal democracies can have chilling effects on their citizens, while totalitarian states actively exploit chilling effects to control their citizens (Macnish [63], 35–37; Zizek [105], 135; [30]). The mere threat of the surveillance of communications can therefore prevent the activities of free speech/expression/association occurring, which is a detriment to the stability of liberal democracies and a mainstay of the stability of totalitarian regimes.

Through reframing the debate away from privacy versus security to that which privacy protects (security) versus security, the equation becomes less straightforward than it may at first have appeared. While privacy is a *pro tanto* right, the right to human security is absolute. This moves from a debate about competing values (privacy or security) to one about the same value. Through providing a common denominator we can now move forward to a proportionality calculation regarding E2EE and terrorism.

## 6 Proportionality

Considerations of proportionality as an element of moral philosophy date back at least as far as Aristotle's *Nicomachean Ethics* [2, bk. V], and in general consideration to the biblical stipulation of "if there is serious injury, you are to take life for life, eye for eye, tooth for tooth, hand for hand, foot for foot" (Exodus 21: 23–24). In contemporary philosophy proportionality has been posited as a central component in the ethics of surveillance [40, 60, 82, 92], intelligence [6, 61, 70, 75], self-defence [95], and jurisprudential sentencing [86]. It also features in three aspects of the typical just war formulation: *jus ad bellum*, *jus in bello*, and the doctrine of double effect (DDE). In the form of DDE proportionality also enters discussions in medical ethics and any other area in which there are foreseeable but unintended harms (see, for example [10, 50, 76]). Appeal has been made to proportionality in writings as diverse as on the environment [88, 94], income distribution [9], investor interests [53], animal welfare [11], and computing [48]. It also sits behind everyday morality as we teach children how to respond to playground taunts and as we determine how to respond to neighbours who regularly play their music too loudly for our comfort.

Proportionality has also become a key consideration in laws regulating surveillance. In the UK, the Regulation of Investigatory Powers Act [77] required acts of surveillance to be both necessary and proportionate [Section 28(2)], as does its replacement, the Investigatory Powers Act [47] [Section 61 (1c)]. In the US, the response of the Supreme Court in the case of *Terry vs Ohio* [392 U.S. 1 (1968)] elicited much discussion as to whether stop and search, within the context of the Fourth Amendment forbidding arbitrary search and seizure, merited a proportionate justification (see the discussion in [86], 1066–70). Proportionality is also a key consideration at the European level of both surveillance practices in general and counterterrorism in particular [16, 69].

Given this considerable breadth and history, proportionality has received comparatively little attention from analytic philosophers. The most notable exception is



Thomas Hurka's *Proportionality in the Morality of War* [46], from which my own *Eye for an Eye: Proportionality and Surveillance* [60] draws to clarify the role of proportionality in surveillance practice (although this approach has been challenged by [92], see also Rønn and Lippert-Rasmussen [82]). More recently still, Henschke has conducted an analysis of proportionality arguments used in the surveillance of metadata debate [40].

Henschke takes his analysis of proportionality beyond the appeal to fairness through balancing harms and benefits to suggest five ways in which the term is frequently used. The first of these, appropriateness, compares the means used with the end sought. In this case, an excessive means to achieve a given end (e.g. using armed police to break up a peaceful demonstration) is disproportionate. The second, action versus inaction, contrasts the act in question with not acting at all. If not acting would be relatively harmless then acting in a way that is harmful would be disproportionate. The third approach is that of comparing costs and benefits. In this case, the costs of doing an action are weighed against the benefits of that action. Fourthly, proportionality may be considered in terms of comparing alternative means to achieving a desired end. The least harmful means would then be the most proportionate, whereas any alternative would be disproportionate. Finally, Henschke suggests a fifth approach which compares simple with complex acts. A simple act (e.g. using armed police to break up a peaceful demonstration) may be excessive, but imagine that a group of known terrorists are using the cover of the demonstration to get sufficiently close to a civilian target with the aim of taking a number of hostages, and there was no alternative means of preventing this from happening. In that case, the use of armed police may not be disproportionate.

In the case of E2EE, the question of proportionality seems at first glance to fit the second approach of Henschke's analysis more cleanly than any of the others. In this way the question can be framed as to whether leaving E2EE in place for common usage (i.e. doing nothing) will promote or enable terrorism over against banning or placing backdoors in E2EE systems. It seems that it would enable positive terrorist activities. Hence removing E2EE seems proportionate according to this approach. However, this conclusion would be too quick. There are also costs to banning E2EE systems, most notably in terms of removing communications security from those who justifiably seek it, such as dissidents persecuted by totalitarian regimes.

Hence the proportionality debate extends beyond a mere contrast between doing nothing and doing something (Henschke's second formulation) to a comparison of costs and benefits (Henschke's third formulation). The benefits of banning E2EE are increased national security in the face of terrorist threats. The costs are decreased personal security for those legitimate users of E2EE, some of whom may face very grave costs indeed. Within liberal democracies, there is a tendency among the majority to feel secure from their own government as a matter of course, thanks to constitutionally robust forms of accountability. As such, the threat of diminished security may not appear particularly great. However, that accountability may be less robust to some members of society (such as minorities in countries experiencing institutional racism in the police) than others. Furthermore, as I shall argue below, the global nature of digital communications in the internet age is such that the costs and



benefits of E2EE must be seen to extend beyond the limits of the (liberal democratic) nation state.

## 7 Maintaining Perspective

It is important to recognize that some level of balance needs to be achieved in a liberal democratic society. While no-one wants a genuinely anarchic, Hobbesian state of nature to prevail in which the malicious can act on their own caprice, nor do citizens in liberal democracies seek to overthrow their governments in favour of totalitarian regimes that promise total security. There is a balance at the heart of liberal democracy that guarantees freedoms to all in order to prevent government abuse, but with the accompanying recognition that some (such as terrorists) will abuse those freedoms to undermine the very state that guarantees those freedoms.

I have suggested above that the rights in question in the E2EE debate are primarily different perspectives on security, but the very fact that they are *perspectives* is central. Take, for instance, the different perspectives of different people groups attempting to take a flight in the post-9/11 world. The perspectives of a person going through security to board a flight are likely to be radically different depending on the ethnicity and apparent religious beliefs of that person [51].

The importance of perspective has been picked up in the ethics of security literature by Jonathan Herington, who has identified three approaches to security: objective, subjective and affective [43]. Taking these in turn, objective security refers to the actual threat a person is under, irrespective of whether they are aware of the threat. For example, I may be about to walk over London Bridge, unaware that several men are planning to drive into people on the pavement and start stabbing them. In this case, I lack objective security, irrespective of my beliefs and feelings. Subjective security refers to the threat a person is aware of. Like objective security, it is a cognitive function: it is something which can be rationalised and argued for. I may be aware that there is a terrorist threat affecting my country. However, while I am aware of the threat (diminished subjective security) I am nowhere near any of the places planned for an attack, and so my objective security remains unaltered. Affective security, by contrast, is an emotive response concerning how secure a person *feels*, regardless of any facts of the matter. In this case, even though I may not be going near a place planned for attack, I may fear that I will be attacked. Equally, I may be going to a place which is the planned location for an attack but, like most in that situation, blissfully unaware of the tragedy that is about to unfold. Each of these is clearly important and each will lead to the individual (or group, for that matter) acting in different ways.

In the aforementioned case of ethnic profiling on flights, statistics clearly point to the fact that non-white members of ethnic minorities have been up to 42 times more likely to be subject to security searches than white people, and are hence less secure than white people in terms of objective, subjective and affective aspects [12, 51, 93]. Notably, such discrimination has been shown to push some members of

ethnic minorities towards terrorism [72, 100]. Furthermore, even when such profiling has taken place, whether based on ethnicity or behaviour, it has been notoriously unsuccessful at identifying terrorists [59], and so its impact on objective security is negligible while the impact on subjective and affective security of those most affected is damaging to the extent of being counter-productive.

The consideration of differing perspectives is not motivated by a desire to see one person or groups perspective diminished as a “mere” perspective (as opposed to fact) but rather to recognize that in talking about balancing rights and interests, a key consideration is “whose rights” are under consideration. We have already encountered this above in considering the rights of citizens in liberal democracies versus those of citizens in totalitarian states. While I have argued that core human rights are important to people in both liberal democracies and totalitarian states, the reason for this importance is different accordingly. In a liberal democracy, society is free because those rights are recognised and upheld. Removing those rights threatens the ongoing stability of the nature of the state. By contrast, in a totalitarian state, dissidents have those rights in name only. They are typically persecuted for enacting those rights if they are caught doing so. Hence E2EE becomes a means of protecting the rights of citizens in liberal democracies and enabling or promoting the rights of citizens in totalitarian states.<sup>4</sup> While many considerations of proportionality are complicated through competing claims in different areas of rights,<sup>5</sup> in this case, the concern is more a matter of how to balance competing concerns regarding the right to security, albeit security from a number of different threats and understood from a number of different perspectives.

We should therefore approach the E2EE debate from (at least) two perspectives: the liberal democratic (domestic) and totalitarian (international).<sup>6</sup> Within the domestic perspective, there is a further differentiation that should be made between the perspective of those in the majority and those in minority groups. Objectively

---

<sup>4</sup> The decision matrix introduced by Jonathan [103] and employed by myself (2016b, 11) in considerations of threats to privacy and security is relevant here. If the surveillant has an interest in the surveillance going ahead and the wrongs of that surveillance will be visited on another, then the surveillant will likely be risk prone: in taking the risk, he suffers no loss (that falls to someone else). Where this is the case, redress can be found through imposing losses on the surveillant. This may arise through the imposition of fines on the surveillant, for example. In this way, when deciding whether to employ surveillance, the surveillant must factor his own loss into the equation.

<sup>5</sup> In his historical overview of the principle of proportionality, Eric Engle traces the idea of justice as proportionality to Aristotle’s Nichomachean Ethics, where it refers to *ratio*, “the right relationship ... between the state and the citizen, adjudicated by the rule of law” ([26], 4). The concept of proportionality then developed through the medieval period via discussions of self-defence and war until Grotius introduced, “the union of the ancient concept of justice as *ratio*, the medieval concept of proportionate self-defence, and the modern concept of balancing interests” ([26], 5). Engle then goes some way to undoing this Grotian union, distinguishing proportionality (which he defines as means end testing in terms of *inalienable* rights) from balancing (which he sees as being cost benefit analysis regarding *alienable* rights) ([26], 10).

<sup>6</sup> Obviously, there are international liberal democratic regimes as well. However, I take the issues for them to be the same as the domestic liberal democratic regime.

speaking, the threat of terrorism is severe and the overall need for communications to be kept secret from governments that practice self-restraint and are held constitutionally accountable (in practice as well as in principle) low. The *objective* security assessment may change, though, depending on the perspective of the person whose security is considered and on the actual level of accountability of the government in question. The *affective* security assessment, by contrast, will likely be radically different for someone who is in the majority group of that democracy than for someone in a minority group, particularly if that minority has suffered historical (and/or ongoing) injustices. As Henschke has pointed out, “It might be easy for me to say that this is the price I am willing to pay ... when it is not me who is likely to bear the costs of misidentification” [41].

Were the domestic perspective the only perspective under consideration, the proportionality calculation would argue against E2EE, albeit that argument would require effective governmental self-restraint and effective accountability, particularly to minority groups. In such cases, E2EE is not strictly necessary, and to allow it would make life considerably easier for terrorists (as noted above). However, I have argued that the domestic perspective needs to be balanced with the international perspective. When this broader picture is considered, along with the potential to challenge totalitarian states, the balance is harder to determine. Nonetheless, the fact that democratic states are, as a rule, more law-abiding than their totalitarian alternatives suggests that within democracies there are alternative means to address terrorism and unjustified state surveillance [8]. By contrast, while totalitarian states are able to continue unthreatened by the people they oppress, they remain immune to their own laws and offer no alternative to those who seek change. As such, E2EE provides a crucial tool in the hands of those who would see democracy come to their own state and, though that, fight abuse and lawlessness.

## 8 Conclusion

In this chapter I have argued that end-to-end encryption (E2EE) presents a fundamental challenge to liberal democratic governments. Communications over E2EE are not accessible to state security services. Hence when terrorists use E2EE, and I have shown that they do, they do so with security. The ongoing existence of E2EE thus indirectly threatens one of the basic responsibilities of the liberal democratic state: the national security of that state. To regain the upper hand in providing security over and against terrorists, liberal democratic states have attempted to challenge E2EE. However, their attempts to do so have been seen to be flawed and ineffective. Furthermore, these attempts come at the expense of citizen privacy.

Privacy is a core value, I have argued, in protecting individuals from the state. However, it is widely recognized as a *pro tanto* value which may be overridden by national security concerns. As such, arguments in favour of privacy which threaten national security may not be convincing, despite the democratic value of privacy.

I have argued that a stronger approach is to look beyond privacy to that which privacy protects, which in this case is the security of the citizen. In this way, national security is weighed against the collected interests of citizen security. This allows for a proportionality calculation to be made. I suggested that while the relevant proportionality calculation appears to be doing something (banning E2EE) over doing nothing, the correct calculation in fact involves weighing the benefits of acting (banning E2EE) over the costs of losing E2EE.

In the case of constitutionally robust and genuinely accountable liberal democratic societies, I argued that there is a morally legitimate security need for private communication, but that this may be overridden by the threat to national security posed by terrorism. Even so, the perspective one takes on this may change depending on how secure within the state one is and feels. However, when taking an international perspective which includes the citizens of totalitarian states, the value of E2EE is considerable in providing a tool for establishing democracy. Given that we live in a world of global communications, it is both unlikely that E2EE can be stopped and it is desirable that we should not want it to be stopped. Terrorism may be the ultimate price that we have to pay for that freedom of others.

## References

1. Allen AL (1988) *Uneasy access: privacy for women in a free society*. Totowa, N.J: Rowman & Littlefield
2. Aristotle, Barnes J (2004) *The Nicomachean ethics*. Edited by Hugh Tredennick. Translated by JAK Thomson. New Ed edition. Penguin Classics, London, England; New York, N.Y.
3. Baker S (2020) A new twist in the endless debate over end-to-end encryption. Reason.Com (blog). 11 Feb 2020. <https://reason.com/2020/02/11/a-new-twist-in-the-endless-debate-over-end-o-end-encryption/>
4. Bamford J (2009) *The shadow factory: the ultra-secret NSA from 9/11 to the eavesdropping on America: The NSA from 9/11 to the eavesdropping on America*. Random House Inc., New York
5. Bamford J (1982) *The puzzle palace: a report on NSA, America's Most Secret Agency*. 5th THUS. Houghton Mifflin
6. Bellaby RW (2014) *The ethics of intelligence: a new framework*. Routledge, London, New York
7. Benn S (1971) Privacy, freedom, and respect for persons. In: Pennock J, Chapman R (eds) *Nomos XIII: privacy*. Atherton Press, New York
8. Bobbitt P (2009) *Terror and consent: the wars for the twenty-first century*. Penguin, London
9. Cappelen AW, Tungodden B (2017) Fairness and the proportionality principle. *Soc Choice Welfare* 49(3–4):709–719. <https://doi.org/10.1007/s00355-016-1016-6>
10. Cavanaugh TA (2006) *Double-effect reasoning: doing good and avoiding evil*. Oxford studies in theological ethics. Clarendon Press, Oxford
11. Cheyne I, Alder J (2007) Environmental ethics and proportionality: hunting for a balance. *Environ Law Rev* 9(3):171–189. <https://doi.org/10.1350/enlr.2007.9.3.171>
12. Choudhury T, Fenwick H (2011) The impact of counter-terrorism measures on Muslim communities. *Int Rev Law Comput Technol* 25(3):151–181. <https://doi.org/10.1080/13600869.2011.617491>

13. Coker M, Yadron D, Paletta D (2015) Hacker Killed by Drone Was Islamic State's "Secret Weapon". *Wall Street J.* 27 Aug 2015, sec. World. <https://www.wsj.com/articles/hacker-killed-by-drone-was-secret-weapon-1440718560>
14. Cruickshank P (2013) Did NSA leaks help al Qaeda? CNN security blogs (blog). 25 June 2013. <https://security.blogs.cnn.com/2013/06/25/did-nsa-leaks-help-al-qaeda/>
15. Davies D (1997) A brief history of cryptography. *Inf Secur Tech Rep* 2(2):14–17. [https://doi.org/10.1016/S1363-4127\(97\)81323-4](https://doi.org/10.1016/S1363-4127(97)81323-4)
16. De Hert P (2005) Balancing security and liberty within the European human rights framework. A critical reading of the court's case law in the light of surveillance and criminal law enforcement strategies after 9/11 special issue on terrorism. *Utrecht Law Rev* 1(1):68–96
17. Dearden L, Sandhu S (2016) Luton delivery driver found guilty of preparing for UK terror attack. *The Independent.* 1 Apr 2016. <https://www.independent.co.uk/news/uk/crime/junead-khan-luton-delivery-driver-found-guilty-preparing-uk-terror-attack-american-forces-a6963451.html>
18. Dodd V (2011) British airways worker Rajib Karim convicted of terrorist plot. *The Guardian.* 28 Feb 2011, sec. UK news. <https://www.theguardian.com/uk/2011/feb/28/british-airways-bomb-guilty-karim>
19. Dooley JF (2018) *History of cryptography and cryptanalysis: codes, ciphers, and their algorithms*, 1st edn. Springer
20. Doyle T (2009) Privacy and perfect voyeurism. *Ethics Inf Technol* 11:181–189
21. Dubrawsky I, Faircloth J (2007) *Security + study guide*. Syngress
22. Ellis JH (1970) The possibility of secure non-secret digital encryption. *UK Commun Electron Secur Group* 6
23. Ellis JH (1987) The story of non-secret encryption. <https://cryptome.org/jya/ellisdoc.htm>
24. Emery F (1995) *Watergate*. Simon and Schuster
25. Engel C (2001) The European charter of fundamental rights a changed political opportunity structure and its normative consequences. *Eur Law J* 7(2):151–170
26. Engle E (2012) The history of the general principle of proportionality: an overview. *Dartmouth Law J* 10:1
27. Farivar C (2018) Australia passes new law to thwart strong encryption. *Ars Technica.* 12 June 2018. <https://arstechnica.com/tech-policy/2018/12/australia-passes-new-law-to-thwart-strong-encryption/>
28. Francis E (2016) Exclusive: Apple CEO Tim Cook says iPhone-cracking software "equivalent of cancer". *ABC News.* 24 Feb 2016. <https://abcnews.go.com/Technology/exclusive-apple-ceo-tim-cook-iphone-cracking-software/story?id=37173343>
29. Frankel S (2016) This is how ISIS uses the internet. *BuzzFeed News.* 12 May 2016. <https://www.buzzfeednews.com/article/sheerafrenkel/everything-you-ever-wanted-to-know-about-how-isis-uses-the-i>
30. Funder A (2004) *Stasiland: stories from behind the Berlin Wall*. New edition. Granta Books
31. Gallagher S (2019) Barr says the US needs encryption backdoors to prevent "Going Dark." Um, What? *Ars Technica.* 8 Apr 2019. <https://arstechnica.com/tech-policy/2019/08/post-snowden-tech-became-more-secure-but-is-govt-really-at-risk-of-going-dark/>
32. Garfinkel S (1995) *PGP: pretty good privacy*. O'Reilly Media, Inc.
33. Garrow DJ (2015) *The FBI and Martin Luther King, Jr.: from 'Solo' to Memphis*. Open Road Media
34. Gavison R (Nov. 1992) Feminism and the public/private distinction. *Stanford Law Rev.* 45(1):1–45
35. Gavison R (1984) Privacy and the limits of the law. In: Schoeman FD (ed) *Philosophical dimensions of privacy*. Cambridge University Press, Cambridge, pp 346–402
36. Genovese MA (1999) *The watergate crisis*. Greenwood Publishing Group
37. Grabenwarter C (2014) *European convention on human rights*. In: *European convention on human rights*. Nomos Verlagsgesellschaft mbH & Co. KG
38. Graham R (2016) How terrorists use encryption. *CTC Sentinel* 9(6):20–25

39. Graham R (2019) Why we fight for crypto. 28 July 2019. <https://blog.erratasec.com/2019/07/why-we-fight-for-crypto.html>
40. Henschke A (2018) Are the costs of metadata worth it? Conceptualising proportionality and its relation to metadata. In: Baldino D, Crawley R (eds) *Intelligence and the function of government*. Melbourne University Press
41. Henschke A (2019) Information technologies and constructions of perpetrator identities. In: Goldberg Z, Knittel S (eds) *Routledge handbook on perpetrator studies*. Routledge, London
42. Herington J (2012) The concept of security. In: Michael S, Christian E (eds) *Ethical and security aspects of infectious disease control: interdisciplinary perspectives*. Ashgate, pp 7–26
43. Herington J (2015) The concept of security, liberty, fear and the state
44. Hoffman LJ (ed) (1995) *Building in big brother: the cryptographic policy debate*. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4612-2524-9>
45. Holden J (2018) *The mathematics of secrets: cryptography from caesar ciphers to digital encryption*. Illustrated Edition. Princeton University Press
46. Hurka T (2005) Proportionality in the morality of war. *Philos Publ Aff* 33(1):34–66
47. IPA (2016) Investigatory powers act. <http://www.legislation.gov.uk/ukpga/2016/25/contents/enacted>
48. Iachello G, Abowd GD (2005) Privacy and proportionality: adapting legal evaluation techniques to inform design in ubiquitous computing. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, CHI'05. New York, NY, USA, pp 91–100. <https://doi.org/10.1145/1054972.1054986>
49. Kahn D (1997) *The Codebreakers: the comprehensive history of secret communication from ancient times to the internet*, 2nd edn. Scribner, New York
50. Kamm FM (1991) The doctrine of double effect: reflections on theoretical and practical issues. *J Med Philos* 16(5):571–585. <https://doi.org/10.1093/jmp/16.5.571>
51. Khaleeli H (2016) The perils of “flying while Muslim”. *The Guardian*. 8 Aug 2016, sec. World news. <https://www.theguardian.com/world/2016/aug/08/the-perils-of-flying-while-muslim>
52. Koops B-J, Kosta E (2018) Looking for some light through the lens of “cryptowar” history: policy options for law enforcement authorities against “going dark.” *Comput Law Secur Rev* 34(4):890–900. <https://doi.org/10.1016/j.clsr.2018.06.003>
53. Krommendijk J, Morijn J (2009) “Proportional” by what measure(s)? Balancing investor interests and human rights by way of applying the proportionality principle in investor-state arbitration. In: Dupuy P-M, Petersmann E-U, Francioni F (eds) *Human rights in international investment law and arbitration*. Oxford University Press, Oxford, pp 421–55. <https://papers.ssrn.com/abstract=2333550>
54. Lever A (2005) *Feminism, democracy and the right to privacy*. SSRN Scholarly Paper, ID 2559971, Social Science Research Network. <http://papers.ssrn.com/abstract=2559971>
55. Lewis P (2010a) Legal fight over spy cameras in Muslim Suburbs. *The Guardian*. 11 June 2010, sec. UK news. <https://www.theguardian.com/uk/2010/jun/11/project-champion-numberplate-recognition-birmingham>
56. Lewis P (2010b) Birmingham stops camera surveillance in Muslim areas. *The Guardian*. 17 June 2010, sec. World news. <https://www.theguardian.com/uk/2010/jun/17/birmingham-stops-spy-cameras-project>
57. Lewis P, Evans R (2013) *Undercover: the true story of Britain’s secret police*. Faber & Faber
58. Locke JL (2010) *Eavesdropping: an intimate history*. OUP Oxford
59. Macnish K (2012) Unblinking eyes: the ethics of automating surveillance. *Ethics Inf Technol* 14(2):151–167. <https://doi.org/10.1007/s10676-012-9291-0>
60. Macnish K (2015) An eye for an eye: proportionality and surveillance. *Ethical Theory Moral Pract* 18(3):529–548. <https://doi.org/10.1007/s10677-014-9537-5>
61. Macnish K (2016a) Persons, personhood and proportionality: building on a just war approach to intelligence ethics. In: Galliot J, Reed W (eds) *Ethics and the future of spying: technology, national security and intelligence collection*. Routledge
62. Macnish K (2016b) Government surveillance and why defining privacy matters in a post-Snowden world. *J Appl Philos*. <https://doi.org/10.1111/japp.12219>

63. Macnish K (2018) *The ethics of surveillance: an introduction*, 1 edn. Routledge, London, New York
64. Macnish K (2020) Mass surveillance: a private affair? *Moral Philos Polit* 1 (ahead-of-print). <https://doi.org/10.1515/mopp-2019-0025>
65. Macnish K, van der Ham J (2021) Cybersecurity ethics. In: *OUP handbook on digital ethics*. Oxford University Press, Oxford
66. Martin L (2018) Bureau Clergyman: how the FBI colluded with an African American televangelist to destroy Dr. Martin Luther King, Jr. *Relig Am Cult* 28(1):1–51. <https://doi.org/10.1525/rac.2018.28.1.1>
67. McKay Tom (2020) Three GOP senators introduce bill experts say would basically ban end-to-end encryption. *Gizmodo*. 24 June 2020. <https://gizmodo.com/three-gop-senators-introduce-bill-experts-say-would-bas-1844157194>
68. Meinrath SD, Vitka S (2014) Crypto war II. *Crit Stud Media Commun* 31(2):123–128. <https://doi.org/10.1080/15295036.2014.921320>
69. Michaelson C (2010) The proportionality principle, counter-terrorism laws and human rights: a German-Australian comparison. *City Univ Hong Kong Law Rev* 2(1):19–44
70. Omand D (2012) *Securing the state*. Hurst, London
71. Perlroth N (2019) What is end-to-end encryption? Another bull’s-eye on big tech. *The New York Times*, 19 Nov 2019, sec. Technology. <https://www.nytimes.com/2019/11/19/technology/end-to-end-encryption.html>
72. Piazza JA (2012) Types of minority discrimination and terrorism. *Confl Manag Peace Sci* 29(5):521–546. <https://doi.org/10.1177/0738894212456940>
73. Pincher C (1978) *Inside story: a documentary of the pursuit of power*, 1st edn. Sidgwick & Jackson Ltd., London
74. Posner RA (2008) Privacy, surveillance, and law. *Univ Chicago Law Rev* 75(1):245–260
75. Quinlan M (2007) Just intelligence: prolegomena to an ethical theory. *Intell Nat Secur* 22(1):1–13
76. Quinn WS (1989) Actions, intentions, and consequences: the doctrine of double effect. *Philos Public Aff* 18(4):334–351
77. RIPA (2000) Regulation of investigatory powers act. <http://www.legislation.gov.uk/ukpga/2000/23/data.pdf>
78. Rachels J (1975) Why privacy is important. *Philos Public Aff* 4(4):323–333
79. Regan PM (1995) *Legislating privacy: technology, social values, and public policy*. University of North Carolina Press, Chapel Hill
80. Roessler B, Mokrosinska D (eds) (2015) *Social dimensions of privacy: interdisciplinary perspectives*. Cambridge University Press, New York
81. Rudd A (2017) We don’t want to ban encryption, but our inability to see what terrorists are plotting undermines our security. *The Telegraph*, 31 July 2017. <https://www.telegraph.co.uk/news/2017/07/31/dont-want-ban-encryption-inability-see-terrorists-plotting-online/>
82. Rønn KV, Lippert-Rasmussen K (2020) Out of proportion? On surveillance and the proportionality requirement. *Ethical Theor Moral Pract*:1–19
83. Sacerdoti G (2002) The European charter of fundamental rights: from a nation-state Europe to a citizen’s Europe. *Colum J Eur L* 8:37
84. Sanger DE, Perlroth N (2015) F.B.I. Chief Says Texas Gunman used encryption to text overseas terrorist. *The New York Times*. 9 Dec 2015, sec. U.S. <https://www.nytimes.com/2015/12/10/us/politics/fbi-chief-says-texas-gunman-used-encryption-to-text-overseas-terrorist.html>
85. Severson D (2017) The encryption debate in Europe. *Hoover Institution Aegis Paper Series*, no 1702
86. Slobogin C (1998) Let’s not bury terry: a call for rejuvenation of the proportionality principle. *John’s L Rev* 72:1053
87. Solove DJ (2002) Conceptualizing privacy. *Calif Law Rev* 90(4):1087–1155
88. Steel D (2013) Precaution and proportionality: a reply to turner. *Ethics Policy Environ* 16(3):344–348. <https://doi.org/10.1080/21550085.2013.844572>
89. Storm M, Cruickshank P, Lister T (2014) *Agent storm: my life inside al-Qaeda*. Penguin



90. Supreme Court (2003) *Lawrence v. Texas*, 539 U.S. 558 (2003). US Supreme Court
91. Taylor L, Floridi L, van der Sloot B (eds) (2016) *Group privacy: new challenges of data technologies*, 1st edn. (2017). Springer
92. Thomsen FK (2020) The teleological account of proportional surveillance. *Res Publica*:1–29
93. Travis A (2016) David Miranda ruling throws new light on schedule 7 powers. *The Guardian*. 19 Jan 2016, sec. UK news. <https://www.theguardian.com/uk-news/2016/jan/19/david-miranda-ruling-throws-new-light-on-schedule-7-powers>
94. Turner D (2013) Proportionality and the precautionary principle. *Ethics Policy Environ* 16(3):341–343. <https://doi.org/10.1080/21550085.2013.844571>
95. Uniacke S (2011) Proportionality and self-defense. *Law Philos* 30(3):253–272
96. United Nations (1948) The universal declaration of human rights. 10 Dec 1948. <http://www.un.org/en/documents/udhr/index.shtml>
97. Waldron J (2006) Safety and security. *Nebr. Law Rev.* 85
98. Warrick J (2016) The “App of Choice” for Jihadists: ISIS seizes on internet tool to promote terror. *Washington Post*. 23 Dec 2016, sec. National Security. [https://www.washingtonpost.com/world/national-security/the-app-of-choice-for-jihadists-isis-seizes-on-internet-tool-to-promote-terror/2016/12/23/a8c348c0-c861-11e6-85b5-76616a33048d\\_story.html](https://www.washingtonpost.com/world/national-security/the-app-of-choice-for-jihadists-isis-seizes-on-internet-tool-to-promote-terror/2016/12/23/a8c348c0-c861-11e6-85b5-76616a33048d_story.html)
99. Wasserman MA (2015) First amendment limitations on police surveillance: the case of the Muslim surveillance program notes. *New York Univ Law Rev* 90(5):i–1826
100. Welch K (2016) Middle Eastern terrorist stereotypes and anti-terror policy support: the effect of perceived minority threat. *Race Justice* 6(2):117–145. <https://doi.org/10.1177/2153368715590478>
101. Whittaker Z (2018) Five years after Snowden: what changed? *ZDNet*. 6 June 2018. <https://www.zdnet.com/article/edward-snowden-five-years-on-tech-giants-change/>
102. Wilber DQ (2017) Here’s how the FBI tracked down a tech-savvy terrorist recruiter for the Islamic state. *Los Angeles Times*. 13 Apr 2017. <https://www.latimes.com/politics/la-fg-islamic-state-recruiter-20170406-story.html>
103. Wolff J (2010) Five Types of risky situation. *Law Innov Technol* 2(2):151–163. <https://doi.org/10.5235/175799610794046177>
104. Zimmerman PR, Ludlow P (1996) *How PGP works/why do you need PGP*. The MIT Press Cambridge
105. Žižek S (2009) *Violence: six sideways reflections*. First Paperback Edition. Profile Books

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Who Should Regulate Extremist Content Online?



Alastair Reed and Adam Henschke

**Abstract** As liberal democracies grapple with the evolution of online political extremism, in addition to governments, social media and internet infrastructure companies have found themselves making more and more decisions about who gets to use their platforms, and what people say online. This raises the question that this paper explores, who should regulate extremist content online? In doing so the first part of the paper examines the evolution of the increasing role that social media and internet infrastructure companies have come to play in the regulating extremist content online, and the ethical challenges this presents. The second part of the paper explores three ethical challenges: i) the moral legitimacy of private actors, ii) the concentration of power in the hands of a few actors and iii) the lack of separation of powers in the content regulation process by private actors.

## 1 Framing the Problem

As liberal democracies grapple with the evolution of online political extremism, social media companies, and their supporting infrastructure, find themselves making more and more decisions about who gets to use their platforms, and what people say online. Moreover, the decision makers at these companies are increasingly uncomfortable with this power. In a public statement from 2018 Facebook CEO Mark Zuckerberg wrote “As I’ve thought about these content issues, I’ve increasingly come to believe that Facebook should not make so many important decisions about free expression and safety on our own” [94]. The CEO of Cloudflare, a company that supports the infrastructure of the internet, Matthew Prince, stated that as “the CEO of a major Internet infrastructure company”, he could wake up in a bad mood and decide that “someone shouldn’t be allowed on the Internet. No one should have that

---

A. Reed (✉)

Swansea University, Singleton Park, Swansea SA2 8PP, Wales, United Kingdom  
e-mail: [alastair.reed@swansea.ac.uk](mailto:alastair.reed@swansea.ac.uk)

A. Henschke

University of Twente, Enschede, Netherlands

© The Author(s) 2021

A. Henschke et al. (eds.), *Counter-Terrorism, Ethics and Technology*,  
Advanced Sciences and Technologies for Security Applications,  
[https://doi.org/10.1007/978-3-030-90221-6\\_11](https://doi.org/10.1007/978-3-030-90221-6_11)

175

power” [15, 63]. At the same time, governments and politicians question whether that Tech Companies are not doing enough to counter extremist content, hate speech, and misinformation online, and that these companies have a social responsibility to go further in their moderation of online content ([8, 53]; Bishop and Macdonald [81], 143).

The question of how and why social media companies and internet infrastructure came to find themselves the arbiters of online speech stems in part from counter-terrorism efforts, particularly the rise of the so-called Islamic State of Iraq and Syria (ISIS) in the mid 2010s. The issue of regulation of political extremists online is inseparable from the evolution of modern global terrorism. In this chapter, we follow the path of that recent history to highlight three ethical concerns arising from the responses to online political extremism. We then suggest that part of the problem with answering ‘who should regulate extremist content online?’ is that there are different aspects to how that content is being regulated. By reflecting on what sorts of institutions and services are being provided, we can suggest a more nuanced and collaborative approach to the regulation of online content.

Regulating extremist content online has faced two particular challenges. The first challenge is determining what extremist content *is*? [46]. Essentially what type of online content should be restricted and potentially removed online in the fight against extremism? This is not a straightforward task when we have no widely agreed upon definition of extremism or violent extremism. Too zealous an approach risks broadening the net too wide, and unnecessarily or disproportionately restricting individuals’ rights and freedoms, whilst a narrower approach allows the free flowing of all but the most extreme of content. Where to draw the line is a controversial decision with no easy solutions [86]. The second is technical: how to *identify* extremist content online. Given the sheer scale and size of many platforms, looking for extremist content is like looking for the preverbal needle in a haystack.<sup>1</sup>

Though we recognise these points, this chapter looks at a less examined challenge of countering extremism propaganda online. Rather than questions of how this should be done, or what material is relevant, this chapter asks questions of *who* gets to make these decisions and *why*. As has been said, freedom of speech is as much about who gets to decide what is said as what is actually said [34].

## 2 The Status Quo: Regulation and Self-Regulation

The division of responsibilities between formal regulation and self-regulation varies across countries depending on local legal frameworks, regulators, and social norms. However, in most countries there is a legal framework that sets out what content is illegal and should be removed online, such as material from a proscribed terrorist organisation, hate speech, or child pornography. These laws then form the foundation

---

<sup>1</sup> Or, as others have noted, given that they are looking for data within data, it is more like looking for a needle in a pile of needles. See: [56].

on which social media companies base their content moderation on. However, the interpretation and enforcement of these laws are increasingly falling on social media companies themselves. As a result, these companies have developed large bureaucracies to monitor and regulate online speech on their platforms. As Jack Balkin argues, whilst in the past regulation of speech was targeted at the speakers, now governments' regulation is increasingly aimed at internet infrastructure (including social media companies), arguing that "in essence, nation-states attempt to get the privately owned infrastructure to do their work for them" ([5], 2015–16). This has led to a form private governance which now plays a central role in regulating extremist content online [5, 44]. In order to deal with the sheer volume of content uploaded daily the major social media companies have developed extensive technological solutions, often utilising artificial intelligence and machine learning,<sup>2</sup> to identify and remove offending content. In addition companies are increasingly collaborating through industry organisations such as the Global Internet Forum for Counter-Terrorism (GIFCT), for example a shared database of terrorist content [8, 26]. As the main platforms have developed their capabilities, they have become increasingly effective at identifying and removing offending content, as Bishop and MacDonald noted, in the case of the largest platforms—Facebook, YouTube and Twitter "referrals from users, law enforcement and governments are responsible for only a small minority of suspensions and take-downs; the vast majority of violations are detected by technology" ([8], 142).

The extent of responsibilities that legal frameworks place on social media companies varies across jurisdictions. In the United States, for example, first amendment free speech protections limit the legal scope of regulating extremist content online, compared to other liberal democracies in Europe or beyond. In Germany, the Network Enforcement Act (NetzDG) requires social media platforms to remove content that is "manifestly unlawful" within 24 h.<sup>3</sup> As Amélie Heldt notes, "[t]he obligation to remove unlawful content is in itself not problematic, but who gets to decide if user-generated content is "manifestly" unlawful? By delegating this task to social media platforms, the State has factually given the responsibility to decide upon the lawfulness of content to the reviewers in charge of content moderation" ([44], 342). Through seeking to make social media platforms more responsible for regulating the content on their platforms, governments have not only given these companies greater powers, but have effectively granted them the moral authority to develop, interpret, and enforce who says what online.

In practise, the regulation of content on a social media platform is governed by a wider set of community rules and guidelines, typically laid out in a platform's own term of service (ToS). As social media companies have evolved from conduits for hosting content to online communities to something more like traditional media,

---

<sup>2</sup> We note here that the use of artificial intelligence and machine learning in identifying and removing extremist content raises separate ethical challenges which are beyond the scope of this paper, but which the authors have addressed elsewhere. See: [46].

<sup>3</sup> The law applies to social media companies with more than two million users in Germany and allows up to seven days for cases that are less clear ([44], 341).

entertainment providers, and/or critical infrastructure, they have become increasingly involved in governing these communities by setting out in their ToS what content is or is not allowed on the platform. Importantly, as these ToS are largely self-generated, the content that is banned by a platform's ToS, can and regularly does go beyond what is merely unlawful. As platforms set out to govern their online communities, this now means not just upholding the law, but also writing the rules about what is acceptable within the community.

These ToS are often then used by law enforcement to request platforms to remove extremist content. Rather than proceeding down a more traditional legal route, with all the complexities that entails, law enforcement can request platforms to remove extremist content by flagging the content to the platforms, pointing out that it violates the platforms own ToS and therefore they should remove it on these grounds [58]. This is the basis on which the Europol's Internet Referral Unit (IRU) works on, which by itself has no enforcement powers. The IRU explains the implications of its referral process: "Thus the decision and removal of the referred terrorist content is taken by the concerned service provider under their own responsibility in reference to their terms of use" ([28], 6). In this situation, the content's removal is due to the material breaking the platform's own rules, and not because it was necessarily illegal.

This brings us to the question of the paper—*who* has the moral authority to make such decisions about what remains online? The role of regulating extremist content online has increasingly shifted from public authorities to private companies, and in the process, these companies now have significant capacity to decide what is allowed online. This brings into focus both the outsourcing to private companies to interpret, and enforce existing legislation, and also the grey area between content that is removed because it breaks a country's laws, and content that is not illegal but removed by private companies because it breaks their own terms of service. Both of these challenges raise questions about the legitimacy of private companies in playing these roles: as we discuss later, in liberal democracies we generally recognise that the state has some moral authority to make such decisions, but it is less clear if private companies have equivalent moral authority.

In response to public and political pressure, platforms can feel compelled to take further actions, and as a consequence update their ToS to ban from their platform a wider array of extremist content. This wider spectrum of what is perceived as extremist content has continued to evolve as perceptions and understandings of how extremists use the internet have developed. For example, Facebook had long banned white supremacy from its platform as 'hateful' content, however it was not until Facebook updated its rules after the Christchurch attacks, that it also included white nationalism and white separatism [30, 71]. As noted at the start of this chapter, this is a place that these CEOs and private companies neither wanted, nor indeed, expected to be in. Interestingly, this challenge has its genesis in the largely uncontroversial decision that terrorist groups like ISIS should not have free run of social media. Terrorism and the means used to combat it are casting a long shadow over the recent evolution of information and communication technologies.

### 3 Terrorism as a Driver for Deplatforming: From ISIS to Political Extremists

The rise of ISIS and its exploitation of propaganda via social media forced social media companies to overhaul their content moderation responses to tackle this new threat. After initial challenges, the major platforms such as Facebook and Twitter were largely successful in driving ISIS supporters off their platforms ([19], 108). However, in comparison to the challenges to come, the situation with ISIS was unique. First, due to their extreme violence and barbaric practices ISIS were almost universally condemned and a general consensus that ISIS and their propaganda should be confronted. Secondly, ISIS material is usually easily identifiable and clearly branded by the group, making its detection much simpler ([19], 108–9).

As Facebook's approaches to content removal developed, so did the list of organisations and individuals that it banned from its platform. However, the individuals and groups banned by Facebook were not just terrorists, but also were increasingly those that operate in the fringes of more familiar political beliefs [35]. In particular with the rise of new far-right movements the line between extremist and mainstream politics became increasingly blurred.

In March 2018 Facebook took the step of banning the British far-right group "Britain First" from the platform, removing its official home page and the pages of its leaders Paul Golding and Jayda Fransen [29, 49]. Stating that "[w]e do not do this lightly, but they have repeatedly posted content designed to incite animosity and hatred against minority groups, which disqualifies the Pages from our service" [29]. Whilst a fringe political organisation, the group had a large social media presence, with more than 1.8 million followers and 2 million likes on its Facebook page, more than double the amount of the Labour party (the mainstream party with the most likes) [49, 61]. The following year Facebook went further designating the group and its leaders under its new definition of 'dangerous groups and individuals', along with a number of other far-right organisations and individuals [89]. This designation also extended the ban to "[p]osts and other content which expresses praise or support for these figures and groups".<sup>4</sup> Although Britain First had ceased to be a political party a few months before the original ban, it is believed that the policy would apply to the proscribed individuals if they ran for or assumed political office in the future [47].

Following a 2018 ban from Facebook, Britain First launched legal action against the company for 'political discrimination' with the group's leader Paul Golding stating: "For too long now social networks have censored certain political viewpoints and thus interfered with the political process" [62].<sup>5</sup> After Facebook's removal of ads supportive of Britain First posted by a third organisation in January 2019,<sup>6</sup> the group accused the company of "political gerrymandering" [90]. A similar position was taken in January 2019 by Saoradh, a political party representing dissident Irish

<sup>4</sup> Facebook statement quoted in: [47].

<sup>5</sup> The group ultimately decided not to continue with the court case the following year [27].

<sup>6</sup> The Facebook adverts were bought by a page called 'Political Gamers TV' supporting a petition by Britain First to halt the reconstruction of a Mosque See: [40, 90].

republicans after Facebook removed its pages from its platform [4, 54]. As it sought a court order to re-instate its pages, Saoradh’s lawyers explained: “Facebook has now taken to remove what they deem to be unacceptable political messages, that sets a very, very dangerous precedent and it’s an attack, a deliberate attack, on the freedom of expression ... Therefore our clients have no alternative but to seek injunctive relief to compel Facebook to uphold what is a very, very basic principle, the right to a political opinion and the right to expression” [54].

Putting aside the nature of Britain First’s and Saoradh political views and the content of their material (which—given liberal democracy’s commitment to political pluralism and free speech—some may find objectionable), the point the groups were making was clear; by choosing to de-platform them, Facebook was interfering in the political process. By exercising their power over which groups can espouse their political views on Facebook, the company had enormous power over what gets said and by whom. Whilst Britain First and Saoradh were on the fringes of mainstream politics, Facebook has also taken steps to ban organisations, which have ‘one foot in the political mainstream’, such as the Greek far-right political party ‘Golden Dawn’<sup>7</sup> which faced a ban by the platform despite having elected members in both the national and European parliament [35, 79].

In September 2019, Facebook removed the account page of the Italian far-right group CasaPound from its platform, along with the pages of its representatives and supporters, on the grounds that they violated Facebook’s Terms of Service by containing hate speech and content that amounted to incitement of violence [41]. In the court case that followed, Facebook was ordered to re-activate CasaPound’s account page, with the court “setting a penalty of €800.00 for each day of violation” [41]. In the court’s ruling it noted, among other points, that “[T]he exclusion of the applicants from Facebook is in contrast with the right to pluralism... eliminating or strongly compressing the possibility for association... to express its political messages.” [72]. This ruling coheres with the view that constraints on free public expression of political beliefs is antithetical to liberal democratic commitments to free and pluralistic societies.

These cases highlight the complex intersection of competing rights that need to be balanced against each other. On the one hand, even in liberal democracies, it is legitimate to restrict content that constitutes hate speech and incitement of violence (Henschke Forthcoming). On the other hand, however, liberal democracies define themselves in part by reference to the right to free speech and political pluralism.<sup>8</sup> This brings us back to the motivating question: who has the authority to decide what extremist content is and what the appropriate responses should be? Should decisions that potentially impact on the public sphere be made by private companies? The near universal agreement that ISIS should be deplatformed has led us to the situation where (fringe) political parties are losing the capacity for public expression. And, while we might agree that the content, and political beliefs of Golden Dawn and the

<sup>7</sup> In October 2020 the leadership of Golden Dawn were convicted of running a criminal organization [6].

<sup>8</sup> For more on this, see: [75, 77, 82, 85].

like are not only objectionable, and perhaps dangerous for democracies, we are still left with the issue of whether social media companies are the right institutions to make decisions about who gets to speak in the new public squares.

These questions were thrown into sharp relief with the 2020 US presidential campaign and the series of events that ultimately led to US President Donald Trump's suspension from many social media platforms. In the run up to the 2020 US presidential elections saw social media platforms revising and updating their policies and terms of service [7]. Twitter in particular, employed extensive measures noting that on Twitter users find “real-time political conversation, resources, and breaking news. And an essential part of our service is taking action on content that attempts to manipulate, disrupt, or cause confusion about civic processes” [83]. As part of this approach Twitter began fact-checking the President's tweets, placing some behind warnings and labelling others as manipulated media or misleading [17, 84].

In the aftermath of the January 6th storming of the Capitol Building, Twitter, Facebook and most major platforms took steps to suspend or ban the President from their platforms [11, 48]. The decision was criticised by many, whilst others supported the actions taken [48]. The then U.S. Secretary of State Mike Pompeo tweeted “Silencing speech is dangerous. It's un-American. Sadly, this isn't a new tactic of the Left” [57]. At the same time, others highlighted the importance of moderating social media content to prevent misinformation, hate speech, and incitement of violence online, but still voiced unease at the process by which the President was suspended from social media.<sup>9</sup> The German Chancellor Angela Merkel, highlighted the importance of free opinion, and that while noting “[t]his fundamental right can be intervened in, but according to the law and within the framework defined by legislators—not according to a decision by the management of social media platforms” [2].

Putting aside any questions over the grounds for the decisions to suspend Trump from social media or questions over whether this decision was biased or politically motivated or not, we need again to ask, who gets to decide this, and why?

## 4 A Deeper Cut: De-Platforming the Platforms

The examples above have focussed on the control that social media platforms have on the content posted on their sites, and also over which individuals and organisations can post content on their platforms. Another type of online de-platforming that has recently emerged focuses not on removing individuals or organisations from a given platform, but literally removing the platform itself from the internet. Online platforms rely on a whole host of auxiliary services to be able to exist online. If these internet infrastructure service providers decide to remove their services, it can prevent the

---

<sup>9</sup> We note here that the governor of the US state of Florida has signed a bill to ban this sort of deplatforming of political actors [42].



platform itself from operating or operating effectively online.<sup>10</sup> The platforms can effectively be de-platformed.

In the wake of the white supremacist violence at a rally in the US city of Charlottesville in 2017, there was a rise in popular and political pressure for private companies to take further action against extremist material online [14, 33, 64]. Whilst the initial focus had been on social media companies to better regulate the content posted on their platforms, a new front opened up, “[t]hat front lies deeper within the web’s infrastructure, in the realm of web hosts, domain registrars, and various other web services. The companies that provide the back-end services of the web have historically resisted pressure to police the behavior of sites that use them and have mostly avoided the spotlight in controversies over online speech” [64]. This marked a change at the level at which extremist content was ‘regulated’ on the internet, and highlights a wider debate about the role of internet infrastructure companies, that support the workings of the internet, and whether they should remain content neutral. In this view internet infrastructure companies are seen as the plumbing of the internet and should not be making decisions about content ([5, 10]; Balkin [39], 2038).

The first major change came in the wake of Charlottesville, when the domain registry service ‘Go-Daddy’ cancelled the registration of the neo-nazi site Daily Stormer. Daily Stormer briefly transferred to Google domains before being cancelled by the provider, which also banned it from YouTube, relegating the website to the dark web [9, 21, 63, 73]. As said, Cloudflare, a service that provides online infrastructure support, including protection from distributed denial of service (DDoS) attacks, soon followed suit [63, 69]. Whilst believing they made the right choice, the CEO Prince expressed unease about the power that Cloudflare could exert [15, 63]. Prince further highlighted that due to the ease at which online attacks could be orchestrated on websites, websites need the services of a network like Cloudflare. Otherwise, they would be at risk of being kicked off-line by anyone that they offend by their content, in practise allowing a form of vigilante justice to police the internet.<sup>11</sup> Going on to note the growing dependence on a few giant networks to provide these services, he argued that soon being online may mean relying on the services of a “company with a giant network like Cloudflare, Google, Microsoft, Facebook, Amazon, or Alibaba” highlighting that Cloudflare by itself already handles 10% of all internet requests [69].

Following the El Paso mass shooting in 2019, Cloudflare decided to take similar action terminating its contract with controversial online platform 8chan,<sup>12</sup> seen by

---

<sup>10</sup> This includes a spectrum of companies that provide services such as web hosting, domain name registries, security (i.e. DDoS protection), online payment processing, among other services. For more see: ([5, 39]).

<sup>11</sup> Prince notes that the initial demands for Cloudflare to terminate their contract with Daily Stormer came from hackers that wanted Cloudflare to ‘[g]et out of the way’ so they could knock it off line with a DDoS attack [69].

<sup>12</sup> 8chan went offline after Cloudflare and other web infrastructure companies refused to provide it with the services in needed to remain online. However, it remerged 3 months later online as 8kun [16]; For more on 8chan’s struggle to stay online see: [18].



many as an online haven for far-right and other extremist views ([20], 12–14, [93]). Prince noted that the El Paso shooter had apparently been inspired by 8chan and had posted a screed to the platform before the attack [70]. Furthermore, this he noted was not an isolated incident, highlighting similar activity earlier that year before both the Christchurch attack on two Mosques and the Poway synagogue attack by lone shooters. Similarly, Prince noted his unease at the arbitrary power the company had, writing “Cloudflare is not a government. While we’ve been successful as a company, that does not give us the political legitimacy to make determinations on what content is good and bad. Nor should it” [70].

These ethical challenges were highlighted again in the wake of the January 6th storming of the US Capitol Building in a series of events that forced ‘free speech’ social network Parler off-line. Positioning itself as a free speech alternative to social media platforms such as Twitter and Facebook, Parler had been one of the fastest growing apps in the preceding months. In the wake of the 2020 presidential election, as platforms like Twitter and Facebook clamped down on misinformation about who had won the elections, millions of conservatives migrated to alternative platforms such as Parler [51].

In the aftermath of the events of January 6th, and former US President Trump’s perceived role in inciting violence numerous social media platforms including Twitter and Facebook took actions suspending the President’s account [11, 48]. As the President and many of his followers sought to migrate to Parler as an alternative platform something unexpected happened. Apple and then Google suspended Parler from their App stores, for not taking sufficient action to police posts made on the Platform. This significantly limited Parler’s ability to gain new followers. Shortly afterwards, Amazon Web Services terminated its contract with the platform for repeated violations of Amazon’s rules in effect taking the platform offline [60]. Parler’s Chief Executive Johan Matze, accused the tech giants of a “coordinated effort” to “completely remove free speech off the internet” [60].

We should not see the actions against Parler in isolation. They were part of wider reactions by private companies in the aftermath of Jan 6th to try to moderate far-right content, as well as mis/disinformation and conspiracy theories on their platforms. However, whilst companies such as Facebook and Twitter took action against content on their own platforms, what makes the case of Parler different, was that Apple, Google and Amazon, took actions against another platform for the content it hosted and a perceived failure to take sufficient action against extremist speech and actions.

Setting aside for one moment questions about the nature of the content on Parler, and whether Parler had or had not taken sufficient action, we have a situation where in effect, a small group of private companies through their actions de-platformed a social media platform over the content it hosted. Given the concentration of power, these companies’ decisions were not just whether to keep Parler as client, but whether Parler could or should remain on the internet. Again, this raises questions about the legitimacy of these platforms to make such far-reaching decisions.

So, what we have seen in just five years is a slippery slope in action. Originally prompted by widespread agreement that the terrorist group ISIS was using social media in ways that connected directly to their violent and extreme actions and beliefs,

we find ourselves in a situation where a then sitting President of the US has been removed from the most popular social media platforms, and now some platforms themselves are finding it increasingly hard to remain online. However, a slippery slope is defined not just by the fact that we have moved from one condition to another one, but that that subsequent condition is one of significant ethical concern [78, 92]. For clarity's sake, reference to a slippery slope is frequently considered to either be a weak argument, or a criticism of a particular argument. This is not our view here—some slippery slopes are of legitimate concern, but one has to be able to show that a given slide from one state to another is occurring, and that the outcome is one that is morally problematic.

The first half of this chapter has shown that we have slid from banning ISIS to deplatforming platforms. The second half of this chapter looks at the ethical concerns with such a slide. In particular, the ethical question that this chapter is now concerned with is whether the private companies ought to be restricting what people say online.

## 5 Ethical Challenges

Again, free speech debates are not so much about what is said, but about who has the authority to decide what is said [34]. In this section we examine three ethical challenges presented by the role private companies play in policing extremist content online. The first is the question of whether these private companies are legitimate actors. The second is the implications of the concentration of power into the hands of a few private companies has on their role of policing the internet. Finally, we look at questions about the lack of separation of power in regulating content online, where private companies become judge, jury and executioner.

As context, let us recall that these issues have largely evolved because governments were reluctant to make decisions about people's political beliefs and their right to public communication. In liberal democracies, censorship, where the state makes the determination on who gets to say what and where, is typically limited to public communications that are highly offensive, are likely to induce or incite significant danger or illegal activity, or that occur in a context of significant and long running discrimination (van Mill [85]; Henschke Forthcoming). One significant reason for this disinclination for governments to decide who gets to say what and where, is that interference in free speech is frequently seen as a marker of authoritarianism. "The right to free speech is hardly in tension with democracy; it is a precondition for it" ([82], 121). When considering the centrality of information and communication technologies to modern life, and the deep integration of social media into people's personal and social lives, we ought to be ethically and politically concerned if governments started making decisions about who says what online.

Whilst in this chapter we highlight some of the ethical challenges of private companies having the power to control content online, it is worth reflecting what happens when this power rest with governments. We have often seen with more authoritarian governments, the blocking of access to social media sites in the face of

criticism and/or in the wake of anti-government demonstrations.<sup>13</sup> So, there is a case for the state to be minimally involved in these decisions. However, as we will show, simply expecting Tech Companies to fill this void poses significant challenges.

### 5.1 *Moral Legitimacy of Private Actors*

The first question that arises is: are private companies legitimate actors to make such decisions? The challenge here is to determine where private companies derive their ‘moral authority’ from to be able to ban content from their platform. In the case of governments an argument based on the social contract could be made, in that “the first duty of government is the security of its people” [50]. The idea is that members of the public cede certain rights to the government and in return the government has a responsibility to provide security.<sup>14</sup> And removing extremist content from social media is a part of this responsibility. However, private companies are not the government. So where do they get their moral legitimacy from?

We suggest that the moral legitimacy of private actors is derived from social licence and responsibility. On social licence, the idea is that companies have a ‘social licence’ “as a means of pursuing new relationships between industry and communities to reflect public values and ensure community support for projects” ([1], 3). These “societal concerns oblige large corporations to act more “responsibly” ... Companies—and their operations—must increasingly satisfy not only the conditions of their formal licences, but also the concerns of host communities and broader society... Hence, it is commonly contended that companies need a “social licence” in addition to their legal and regulatory obligations” ([66], 341). Much like the social contract that allows a government to make decisions about the lives of its citizens, the social licence afforded private companies is something that society, or at least relevant members in that society, grant to that company. “Typically, an operation’s social licence is theorised as comprising ongoing acceptance or approval from the local community” ([66], 344). Their moral legitimacy comes in part from the agreement of society that they can continue to exist, in order to provide goods and services and so on.

Parallel to this is the notion of responsibility, whereby the company can be held responsible for the outcomes of their decisions. If a company is shown to be causing undue harm to the environment, or to people, they may be held responsible for that. Similarly, if they are shown to *act responsibly*, by admitting to, and responding to negative impacts of their practices, then the community may see them as earning legitimacy to operate. “Moral legitimacy can be achieved by engaging with affected

---

<sup>13</sup> For example see: [24, 55].

<sup>14</sup> We note here that this opens a larger discussion of the ethics of state use of power, legitimacy, and the tension between state-sponsored censorship and the responsibility of the state to provide security to its citizens. There is not space in this chapter to cover those questions, however. For see more on this see: [22, 23, 76].

persons and groups and by finding solutions and compromises with them in order to overcome dissent” ([25], 679). The point here is that the ethical legitimacy is not simply earned, but must be attended to and maintained. Arguably, when the threats posed by ISIS’ online activities became more apparent, the social media companies had to act in a way that showed that they were responsible—they saw the role that their services and products played in advancing ISIS’ activities, and acted in ways that showed, or at least purported to show, that they cared about the negative impacts that their services were allowing. This brings to the surface a deeper point. The question of who makes decisions, leads to a second question of *how* they make the decisions. The moral legitimacy of the actor in part depends on how they make decisions.

This does not mean that the answers are necessarily clear or easy. For instance, like any issues of representation, who counts as society? What happens if one significant sector of society deeply believe that a political actor needs to be deplatformed, while another significant sector of society deeply believe that that political actor represents their views, and so needs to retain their right to public communication? Moreover, how do we know when society has withdrawn that social license? The purpose of this chapter is not to offer answers to these questions, but instead it is to first show that in the vacuum left by governments reluctant to censor political extremists, Tech Companies are the pivotal actors. Second, we hope to show the contours of where their moral authority to make relevant decisions might come from. Finally, we are pointing out particular emerging questions which would form the basis for discussions moving forward.

## 5.2 *Concentration of Power*

In the online world, power is concentrated into the hands of a few giant private tech companies (Fernandez [32]; Kang [37, 52], 2). In terms of social media, the field is dominated by a few big players Facebook (including Instagram and Whatsapp), YouTube and Twitter (Statista n.d.; [13], 88–92).<sup>15</sup> For web infrastructure companies the picture is less clear as there is a much wider diversity of small and large companies across the plethora of infrastructure services. However, as the examples with Daily Stormer, 8chan and Parler show, there are limited options, with the technical ability and capacity, to keep platforms online at scale. This concentration of power into a few hands, as we argue below, should necessitate a higher level of scrutiny and new obligations on the decision makers. Through executing their power to decide who is or is not allowed on their platforms, a few private companies are in practice deciding who can have a voice online.

If there was a much larger plurality of social media platforms, then the banned individual or group could simply move to another platform. Having a plurality of

---

<sup>15</sup> Further, Evelyn Douek has argued that the increased collaboration of platforms through industry bodies to confront extremist content can “augment the power of already powerful actors by allowing them to decide standards for smaller players” [26].

the press means that a small amount of decision makers do not get to determine what is or is not the news, or whose ideas are allowed in political debate. With this concentration of power of both social media, and the supporting infrastructure, liberal democracies are at risk of significantly narrowing the set of people who get to make the decisions about the public communication of political ideas. Note also that this is a global phenomenon—decisions by Facebook, Twitter, Google etc., impact not just the national discourse, but discussions everywhere. The point here is that, in liberal democracies committed to political pluralism, we need not just multiple avenues for public expression of ideas, but for these avenues to encompass a range of views.

The concentration of power with a small set of companies means that a small group of tech executives get to decide who can and who cannot have a voice online. Such power we argue, and its far-reaching implications on public debate, necessitates that these decisions should face a far higher level of scrutiny than they currently do. One of the main concerns about government censorship is that the concentration of power necessitated by government allows for a very small number of people to make decisions that impact a large number of people. The evolution of the internet, the rise of a small set of companies to dominate the social media space, and the dependence of them and other smaller companies on an increasingly small number of service providers means that we are seeing a similar concentration of power in the hands of a few key actors. To be clear, we are not necessarily disagreeing with decisions to deplatform ISIS or other political extremists. Rather, our point is around those with decision making power in these private companies have power that is significantly disproportionate to those who are affected by their decisions. To explain the concern here, in liberal democracies, the authority of the state lies in people having the ability to vote in elections to bring about a peaceful transitions of power. Private companies have none of these, and given the market dominance of a small number of companies, people have very limited ability to choose other platforms. And in 4 years there is no election to decide who the next Twitter should be.

A final aspect to this concentration of power is that many of these companies, whether the public facing or the supporting infrastructure, are effectively US based. The issue here is the values and standards that are being developed, interpreted, and applied, have not just a developed world view on things, but will typically have an English speaking and American framing and foundation to these issues ([44], 340). Moreover, the social and legal factors that will influence the decisions about whether to protect or deplatform a speaker or company will be heavily US based. If, for instance, there are significant protests in Fiji about particular content, or particular views being deplatformed, is this going to have the same effect as significant protests in the US? Moreover, in line with the point above, if people are dissatisfied with particular laws or ToS in the US, they can seek to change those laws through political processes. But if people in Fiji are dissatisfied, then what processes are available to them to change US laws or ToS? The point here is that, not only is there a concentration of decision making power, but also a concentration of social power.

The overall point here is that, given the way that social media and supporting services have evolved, we are facing an issue now that a very limited number of

people have the capacity to decide who gets deplatformed, and who doesn't. Again, the question we need to ask is, is this concentration of power ethically, socially, and politically justifiable? We suggest here that the answers to this are going to require not just a consideration of why these decision makers have this moral authority, if at all, but will also require reflection on what sorts of institutions these private companies are (see below), as well as further reflection on the best ways to ensure decision making occurs in a way that is responsive to a range of stakeholders' views and concerns, a point we return to at the end of the chapter.

### 5.3 *Lack of Separation of Power*

The third area of concern is the lack of separation of powers. This builds on the point made above, that an important part of question of who makes the decisions, is how they make decisions. In terms of online content regulation, as the whole process is in effect carried out by the private companies themselves, private companies are in practice acting as prosecutor, judge, jury and executioner.<sup>16</sup> The platforms, through their ToS, determine the rules for governing the content allowed on their platforms, with the interpretation and enforcement of these rules at the companies' discretion [44]. The sanctions invoked in the case of breaking the terms of service range from removing content, to suspending accounts to banning individuals or organizations from the platform are similarly decided by the platform. And finally, appeals processes are run by the platforms themselves. As a result, this private governance of online speech raises questions of transparency and due process ([5], 2031). In short questions about how they make decisions, and if they make them in the appropriate way.

The argument here is not of any impropriety on behalf of the private companies in enforcing their terms of service,<sup>17</sup> rather that separation of powers is a well-established safeguard in liberal democracies against potential abuses of power. In most liberal democracies there is some version of separation of powers in government, between the legislature, judiciary and executive.<sup>18</sup> The fact that private companies currently decide the rules of what is allowed to be said online, enforce these rules, and run the appeals processes concentrates even more power in their hands and removes potential safeguards against abuse or bias. The ethical issues here obviously arise in situations where that concentrated power is abused—if a tech company capriciously decides what the rules of online activity are, who says what, and/or how any appeals are run, we have a system that lacks the basic pre-conditions of justice. There is a further issue of bias. We find expression of this in discussions of criminal

---

<sup>16</sup> Balkin highlights this as the problem of 'collateral censorship' that emerges from private governance of online regulation ([5], 2031).

<sup>17</sup> It should be noted that many tech companies have been taking steps to increase the transparency of their content moderation practices, and to include clear appeals processes.

<sup>18</sup> For more on this, see for example: [67, 87, 88].

justice and the separation of powers. Christopher Wellman notes: “The crucial point is that things would deteriorate into a horribly dangerous mess if each individual were personally responsible for punishing those who wronged her. The explanation for this has been laid out plainly by social contract theorists... victims who personally mete out the punishment are more likely to punish the innocent and over-punish the guilty” ([91], 428). Wellman’s point is that the ends of justice are better met when there is a separation of powers. We suggest that a similar principle likely arises here—the decisions about who says what online are better made if there is some effective separation between those who write the rules, enforce them, and then act to adjudicate disputes about their enforcement.

However, it should be noted that these concerns are not lost on the companies. Facebook for example has recently created an independent ‘Oversight Board’ as it believed “that it shouldn’t be making so many decisions about speech and online safety on its own” [31]. The board’s role is to “review a select number of highly emblematic cases and determine if decisions were made in accordance with Facebook’s stated values and policies” [31]. The decisions made by the board will then be binding for Facebook to implement (unless it breaks the law) [31]. This board had its first major test following Donald Trump’s statements about the January 6th insurrection at the US Capitol Building, and decided that—due to the risks he posed to public safety, the former US President would remain banned from Facebook until 2023 [74]. Heldt has argued that through the establishing of an independent oversight board and the publication of the guidelines by which its moderators interpret the rules in Facebook’s ToS, it has created “structures and procedures similar to administrative law” ([44], 354). Whilst Facebook has taken steps to add in elements of independence and separation of power into their bureaucracies of private governance, questions remain about the legitimacy of institutions like Facebook’s Oversight Board, when compared to the separation of powers within a liberal democracy.

So, now we can see that Facebook is attempting to develop a set of processes that divest the company of particular decisions about who gets deplatformed, when, why, and for how long. But this raises further questions—first, how independent is this Oversight Board? Whoever gets to decide the make-up of the board has significant power to influence the direction that any future decisions are made. Second, even if the board itself acts fairly independently, what happens when a decision by the board is likely to have significant economic costs? How does Facebook management adjudicate between the board’s decisions and the economic advice? Finally, this gives Facebook a significant advantage over other would-be social media companies. How can a smaller company, much less a start-up have the capitol to setup this quasi-legal infrastructure from the outset that now society deems important? This brings us back to the issue of concentration of power, discussed earlier. Again, our point is not to offer answers to these questions, but to map out the complexity faced when deciding to deplatform political actors. Where it was once relatively easy to make a decision to remove ISIS’ material, we are now faced with a highly complex space.



## 6 Different Institutions, Different Ethical Responsibilities

We have discussed the notion of private companies having a social licence to operate, and that they may lose their moral legitimacy if they do not act responsibly. A further point about legitimacy derives from what type of institution we are discussing.<sup>19</sup> In this section we examine whether we can see social media companies, and the companies that provide the supporting infrastructure, simply as private companies, or as news media companies, or are they instead public infrastructure? Different answers to these questions suggest different ethical responsibilities. Our suggestion here is that not only can we see different social media and related companies differently, we ought actually to see them differently depending on the service that they provide.

If we consider social media and web infrastructure companies to simply be private companies, according to a view like that of shareholder theory,<sup>20</sup> their obligations are restricted to following the law and providing the best returns to shareholders [38]. If seen as purely private companies, they would have on the one hand no obligation to remove or restrict extremist content online beyond what is purely illegal. In short, unless the individual or group is proscribed and/or the material breaks relevant laws such as incitement to violence or hate crimes, the company would have no obligation to remove or restrict access to the content. On the other hand, as a private company, private companies are free to set their own terms of service, deciding both what can be said on their platforms and by whom. Hence, they have no obligation to provide access to their platform to everyone. Thus, seeing them simply as private companies, as long as no laws are broken, then they have no particular responsibility to regulate what is online. However, at the same time, just as any other private company can refuse to offer a product or service to a client, then these companies are free to regulate online content as they will.

Instead, we could consider social media platforms as *media* companies, with all the editorial oversight requirements and responsibility for the content published that this requires. In which case social media platforms might be legally liable for all of the content posted on their platforms. In contrast to the current situation in which platforms are not held liable for the content of their users, a legacy of the ‘safe harbor’ provisions in section 230 of the US 1996 Communications Act.<sup>21</sup> These provisions have underpinned much of the evolution of social media platforms, and we note here that change in such a position would challenge the concept of social media as we currently know it, which has been built on the premise that they are not responsible for the content of their users’ posts.

Alternatively, it could be argued that private companies are providing a public good, and should be seen as a provider of a *public infrastructure utility*, like a water or electricity company. This different conception prompts us to consider that their

---

<sup>19</sup> For more on this, see: [59].

<sup>20</sup> We note here that a theory like stakeholder theory might take a different view, that the tech companies have a broader set of commitments that includes people beyond shareholders. For more on stakeholder theory see: [68].

<sup>21</sup> For a wider discussion on the debate see: [36].



moral responsibilities might be different than a normal public company. “If they are, instead, more like public infrastructure—like that of a road system or energy system—then they may have to constrain their responsibility to shareholders and profits by reference to public safety and extremist content that poses a public safety threat would likely be justifiably disrupted” [46]. In providing a public good such institutions have a responsibility to the *safety* of their users. In the case of social media companies and web infrastructure companies this responsibility could be seen to include keeping their users safe from extremist content online. This responsibility for safety offers the justification for their content moderation.

However, seeing private companies as public utilities likely generates wider obligations. Public utilities are normally heavily regulated and required to provide equal access to their services to everyone. For example, water or electricity utilities are usually required to provide their services to all members of the public and not discriminate in their choice of customers.<sup>22</sup> In this case if we see private companies as public utilities, this likely places wider obligations on them to provide equal access to their platforms or services.

This leads to another question, should companies that provide different types of service be seen as different types of institution, and hence have different responsibilities and obligations? For example, should social media companies such as Facebook and YouTube, which host, organise and promote users’ content, be seen differently to internet infrastructure companies such as Cloudflare which support the back-end of the web?<sup>23</sup> This point highlights the ongoing debate about the appropriate level at which content on the internet should be regulated. However, this is not an argument for the latter to have no responsibility to regulate. Rather, it is to say that all should have some level of responsibility for regulating the content that their services support. For example, enforcing action against material from proscribed terrorist groups, or other illegal content such as child pornography. However, should we expect the same regulation of content by social media companies, as by content delivery networks, web hosts or domain registrars? [64]. And if so, is this best understood by seeing them as different types of institutions, entailing different levels of responsibilities and obligations?<sup>24</sup>

For example Balkin argues, “[d]ifferent parts of the internet infrastructure<sup>25</sup> should have different responsibilities to protect freedom of speech online” ([5], 2037). He sets out three groups of companies with different responsibilities: Basic Internet Services (including hosting services, telecommunications services, domain name

---

<sup>22</sup> For a wider discussion on whether internet infrastructure companies should be seen as delivering a public good see: [39].

<sup>23</sup> This is an argument made by Cloudflare’s Mathew Prince [70]. Suzanne van Geuns and Corinne Cath-Speth, put forward an argument that web infrastructure companies like Cloudflare should really be seen as traffic controllers, highlighting the choices that these companies make over the flow of traffic on the internet [39]. However, we also note that the distinction is increasingly blurred with some big tech companies including both social media platforms and web infrastructure services.

<sup>24</sup> We note here that there is a wider debate here about the neutrality of the inner workings of the internet which is beyond the scope of this chapter. For a brief discussion see: [10].

<sup>25</sup> Balkin includes social media companies in his definition of internet infrastructure.

services, caching and defense services), Payment Services and Content Curators (including social media and search engines). Basic Internet Services (such as Cloudflare) and Payment services he argues should not regulate content, while content curators like social media companies have different responsibilities ([5], 2038). Whilst other would argue that basic internet service companies should still have some responsibility to regulate content (as noted above), it might still be reasonable to expect that these responsibilities are different to and less extensive to those of social media companies. If we see companies as having different levels of responsibility depending on the service they provide, this maybe best understood as seeing them as different types of institutions which determines their ethical responsibilities and their legitimacy to take action to regulate.

## 7 Conclusion: Is Co-Regulation a Solution?

We have argued in this paper that the role that social media companies and web infrastructure companies play in regulating extremist content online, raises significant ethical questions that warrant further attention. Similarly, we have noted above that placing responsibility for online regulation and its enforcement solely within the remit of governments also causes significant ethical dilemmas, in particular with authoritarian states. One approach to resolving these challenges is the idea of co-regulation. In the current situation, internet regulation is partially covered by both public authorities and also self-regulation by social media and web infrastructure companies. The ethical dilemmas highlighted in this chapter fall within the grey area between public authorities' regulation and companies' self-regulation, one solution is to move towards a more dynamic regulatory environment between both actors in which regulatory frameworks are made in a consensual manner. However, as Natali Helberger, Jo Pierson and Thomas Poell have argued "the realization of core public values in these sectors should be the result of the dynamic interaction between platforms, users, and public institutions" ([43], 10).

A first benefit of co-regulation is that it can help reduce the problems of granting authority to one single powerful sort of actor. "These decisions, which deeply affect public safety, the character of public communication, and freedom of expression, should not be left to governments, or to individual platforms and their users. As history shows over and over again, unilateral government regulation of public communication tends to sit in tension with freedom of speech. Furthermore, since social media corporations are primarily driven by commercial interests, they cannot be trusted to always act in the interest of the public good either" ([43], 8). Ideally, co-regulation is a way of not just balancing different interests, but ensuring that each set of actors limit the other set's interests.

The benefits of a co-regulation approach are that it avoids either effectively letting the state be the sole decision maker about political speech, or the challenges of simply

having companies engage in self-regulation [65]. Whilst the latter suffers from questions of accountability and legitimacy [12] outlined above, public authority regulation not only runs the risk of the state being the arbiter of political speech, but often suffers from being slow and unresponsive and lacking the necessary technological know-how [3]. Through developing a process by which the different actors can work more dynamically together, maybe we are able to go some way to resolving the ethical challenges above. Co-regulation between governments and private companies may increase the democratic accountability of who regulates the internet, and in part answer some of the questions of the moral legitimacy of actors. Furthermore, the closer interaction of governments and private companies, may also provide an avenue to address the challenges of concentration of power and separation of powers. That said, we recognise that such public/private cooperation may not only leave the ethical issues we have identified unresolved, but close collaboration between government and private interests may increase the problems identified. It is beyond the scope of this chapter to specify what co-regulation involves; our point here is that if we're talking about how to deal with extremist material online, co-regulation is an area that has ethical significance. All that said, we consider that this is the opening to a discussion on how best to regulate politically extreme content online, rather than a solution. Co-regulation offers one part of that solution, but needs further discussion.

At the heart of content regulation is balancing of rights, between protecting users from extremist material and upholding freedom speech and free public communication. The moral legitimacy of private companies to restrict rights in the pursuit of safety of its users is dependent on what type of institution they are. Furthermore, the obligations of private companies to protect rights of freedom of speech and communication also depends on what type of institution they are. Hence answering this question is fundamental to determining the role they should play in content regulation. While we have travelled far from shutting down ISIS, it is clear that the need to limit terrorist content online has effectively been the start of discussions about who gets to regulate content online, and these discussions are likely to go on for a long time.

## References

1. Aitken M, Toreini E, Carmichael P, Coopamootoo K, Elliott K, van Moorsel A (2020) Establishing a social licence for financial technology: reflections on the role of the private sector in pursuing ethical data practices. *Big Data Soc* 7(1):2053951720908892. <https://doi.org/10.1177/2053951720908892>
2. AP News (2021) Germany's Merkel: trump's twitter eviction 'problematic.' AP NEWS. 11 Jan 2021, sec. Donald Trump. <https://apnews.com/article/merkel-trump-twitter-problematic-dc9732268493a8ac337e03159f0dc1c9>
3. Ayres I, Braithwaite J (1992) *Responsive regulation: transcending the deregulation debate*. Oxford University Press, USA
4. Bain M (2019) Facebook sued by Dissident Group Saoradh over removed pages. *Belfasttelegraph*. 31 Jan 2019. <https://www.belfasttelegraph.co.uk/news/northern-ireland/facebook-sued-by-dissident-group-saoradh-over-removed-pages-37767707.html>

5. Balkin JM (2018) Free speech is a triangle. *Columbia Law Rev* 118:2011–2056
6. BBC (2020) Greece golden dawn: Neo-Nazi leaders guilty of running crime gang. BBC News. 7 Oct 2020, sec. Europe. <https://www.bbc.com/news/world-europe-54433396>
7. Bergengruen V (2020) ‘The devil will be in the details.’ how social media platforms are bracing for election chaos. *Time*. 23 Sept 2020. <https://time.com/5892347/social-media-platforms-bracing-for-election/>
8. Bishop P, Macdonald S (2019) Terrorist content and the social media ecosystem: the role of regulation. In: *Digital Jihad: online communication and violent extremism*, pp 135–152. <https://cronfa.swan.ac.uk/Record/cronfa52902>
9. Brandom R (2017a) Google says it will ban Neo-Nazi site after domain name switch. *The Verge*. 14 Aug 2017. <https://www.theverge.com/2017/8/14/16145064/google-daily-stormer-ban-neo-nazi-registrar-godaddy>
10. Brandom R (2017b) Charlottesville is reshaping the fight against online hate—the verge. *The Verge*. 15 Aug 2017. <https://www.theverge.com/2017/8/15/16151740/charlottesville-daily-stormer-ban-neo-nazi-facebook-censorship>
11. Byers D (2021) How Facebook and Twitter decided to take down trump’s accounts. *NBC News*. 14 Jan 2021. <https://www.nbcnews.com/tech/tech-news/how-facebook-twitter-decided-take-down-trump-s-accounts-n1254317>
12. Campbell AJ (1999) Self-regulation and the media. *Federal Comm Law J* 51(3). <https://www.repository.law.indiana.edu/fclj/vol51/iss3/11/>
13. Cicilline DN (2020) Investigation of competition in digital markets: subcommittee on antitrust commercial and administrative law of the committee on the judiciary, 449
14. CNBC (2017) Internet firms flex muscle to exile white supremacists. *CNBC*. 16 Aug 2017, sec. Technology. <https://www.cnb.com/2017/08/16/internet-firms-flex-muscle-to-exile-white-supremacists.html>
15. Conger K (2017) Cloudflare CEO on terminating service to Neo-Nazi site: ‘the daily stormer are assholes.’ *Gizmodo*. 16 Aug 2017. <https://gizmodo.com/cloudflare-ceo-on-terminating-service-to-neo-nazi-site-1797915295>
16. Conger K (2019) It’s back: 8chan returns online. *The New York Times*. 4 Nov 2019, sec. Technology. <https://www.nytimes.com/2019/11/04/technology/8chan-returns-8kun.html>
17. Conger K (2020) Twitter has labeled 38% of trump’s tweets since tuesday. *The New York Times*. 5 Nov 2020, sec. Technology. <https://www.nytimes.com/2020/11/05/technology/donald-trump-twitter.html>
18. Conger K and Popper N (2019) Behind the scenes, 8chan scrambles to get back online. *The New York Times*. 5 Aug 2019, sec. Technology. <https://www.nytimes.com/2019/08/05/technology/8chan-website-online.html>
19. Conway M (2020) Routing the extreme right. *RUSI J* 165(1):108–113. <https://doi.org/10.1080/03071847.2020.1727157>
20. Conway M, Macnair L, Scrivens R (2019) Right-wing extremists’ persistent online presence: history and contemporary trends. *ICCT Policy Brief* 24
21. Cox J (2017) After shutdown, daily stormer users are moving to a dark web version of site. 15 Aug 2017. <https://www.vice.com/en/article/evvxvz/white-supremacist-website-daily-stormer-goes-offline>
22. Cudd A (2012) Contractarianism. *Stanford Encyclopedia of Philosophy/Winter 2013 Edition*. <https://plato.stanford.edu/archives/win2013/entries/contractarianism/>
23. D’Agostino F, Gaus G, Thrasher J (2011) Contemporary approaches to the social contract. *Stanford Encyclopedia of Philosophy/Winter 2012 Edition*. <https://plato.stanford.edu/archives/win2012/entries/contractarianism-contemporary/>
24. Deahl D (2018) Iran has banned telegram after claiming the app encourages ‘armed uprisings.’ *The Verge*. 1 May 2018. <https://www.theverge.com/2018/5/1/17306792/telegram-banned-iran-encrypted-messaging-app-russia>
25. Demuijnck G, Fasterling B (2016) The social license to operate. *J Bus Ethics* 136(4):675–685. <https://doi.org/10.1007/s10551-015-2976-7>

26. Douek E (2020) The rise of content cartels. Knight First Amendment Institute, The Tech Giants, Monopoly Power, and Public Discourse. February. <https://knightcolumbia.org/content/the-rise-of-content-cartels>
27. Erwin A (2019) Britain first ends Facebook challenge over Northern Ireland page ban. Belfast-telegraph. 15 May 15. <https://www.belfasttelegraph.co.uk/news/northern-ireland/britain-first-ends-facebook-challenge-over-northern-ireland-page-ban-38114561.html>
28. EUROPOL (2020) EU IRU transparency report 2019. Europol. 13 Oct 2020. <https://www.europol.europa.eu/publications-documents/eu-iru-transparency-report-2019>
29. Facebook (2018) Taking action against Britain first. Taking action against Britain first (blog). 14 Mar 2018. <https://about.fb.com/news/h/taking-action-against-britain-first/>
30. Facebook (2019) Standing against hate. About Facebook (blog). 27 Mar 2019. <https://about.fb.com/news/2019/03/standing-against-hate/>
31. Facebook Oversight Board. n.d. Oversight board independent judgement. Transparency. Legitimacy. Facebook Oversight Board. Accessed 12 June 2021. <https://oversightboard.com/>
32. Fernandez R, Adriaans I, Klinge TJ, Hendrikse R (2021) How big tech is becoming the government. SOMO. 5 Feb 2021. <https://www.somo.nl/how-big-tech-is-becoming-the-government/>
33. Finkle J, Rodriguez S (2017) Tech companies in the crosshairs on white supremacy and free speech. Reuters. 14 Aug 2017. <https://www.reuters.com/article/us-virginia-protests-godaddy-idUSKCN1AU0CV>
34. Fish S (1994) There's no such thing as free speech: and it's a good thing, too. Oxford University Press, New York. <https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=273279>
35. Fisher M (2018) Inside Facebook's secret rulebook for global political speech—The New York Times. The New York Times. 27 Dec 2018. <https://www.nytimes.com/2018/12/27/world/facebook-moderators.html>
36. Flew T, Martin F, Suzor N (2019) Internet regulation as media policy: rethinking the question of digital communication platform governance. *J Digital Media Policy* 10(1):33–50. [https://doi.org/10.1386/jdmp.10.1.33\\_1](https://doi.org/10.1386/jdmp.10.1.33_1)
37. Floridi L (2021) Trump, Parler, and regulating the Infosphere as our commons. *Philosophy and Technology*, March. <https://doi.org/10.1007/s13347-021-00446-7>
38. Friedman M (1972) The social responsibility of business. The New York Times Magazine, 13 Sept. In: Friedman M (ed) *An economist's protest: columns on political economy*. Thomas Horton & Daughters, Glen Ridge, NJ, pp 177–84
39. Geuns SV, Cath-Speth C (2020) How hate speech reveals the invisible politics of internet infrastructure. Brookings (blog). 20 Aug 2020. <https://www.brookings.edu/techstream/how-hate-speech-reveals-the-invisible-politics-of-internet-infrastructure/>
40. Ghosh S (2019) Facebook cleared political ads for a far-right group it banned just 8 months ago. Business Insider Nederland. 8 Jan 2019. <https://www.businessinsider.nl/far-right-group-britain-first-facebook-ads-mosque-ban-2019-1/>
41. Global Freedom of Expression (2020) Facebook v. CasaPound. Global Freedom of Expression. <http://globalfreedomofexpression.columbia.edu/cases/casapound-v-facebook/>
42. Godwin C (2021) Florida governor signs bill to ban big tech 'Deplatforming.' BBC News. 24 May 2021, sec. Technology. <https://www.bbc.com/news/technology-56952435>
43. Helberger N, Pierson J, Poell T (2018) Governing online platforms: from contested to cooperative responsibility. *Inf Soc* 34(1):1–14. <https://doi.org/10.1080/01972243.2017.1391913>
44. Heldt A (2019) Let's meet halfway: sharing new responsibilities in a digital age. *J Inf Policy* 9:336–369. <https://doi.org/10.5325/jinfopoli.9.2019.0336>
45. Henschke A (Forthcoming) Free speech, free public communication and counterterrorism. In: Feltes J, Henschke A, Miller S, Elgar E (eds) *Counter-terrorism: the ethical issues*
46. Henschke A, Reed A (2021) Toward an ethical framework for countering extremist propaganda online. *Stud Conflict Terrorism*. <https://doi.org/10.1080/1057610X.2020.1866744>

47. Hern (2019) Facebook bans far-right groups including BNP, EDL and Britain first. *The Guardian*. 18 Apr 2019, sec. Technology. <http://www.theguardian.com/technology/2019/apr/18/facebook-bans-far-right-groups-including-bnp-edl-and-britain-first>
48. Hern (2021) Opinion divided over trump's ban from social media. *The Guardian*. 11 Jan 2021. <http://www.theguardian.com/us-news/2021/jan/11/opinion-divided-over-trump-being-banned-from-social-media>
49. Hern A, Rawlinson K (2018) Facebook bans Britain first and its leaders. *The Guardian*. 14 Mar 2018, sec. World news. <http://www.theguardian.com/world/2018/mar/14/facebook-bans-britain-first-and-its-leaders>
50. Heyman S (1991) The first duty of government: protection, liberty and the fourteenth amendment. *Duke Law J* 41(3):507–571
51. Isaac M, Browning K (2020) Fact-checked on Facebook and Twitter, conservatives switch their apps. *The New York Times*. 11 Nov 2020, sec. Technology. <https://www.nytimes.com/2020/11/11/technology/parler-rumble-newsmax.html>
52. Kang C, McCabe D (2020) House lawmakers condemn big tech's 'monopoly power' and urge their breakups. *The New York Times*. 6 Oct 2020, sec. Technology. <https://www.nytimes.com/2020/10/06/technology/congress-big-tech-monopoly-power.html>
53. Kayali L, Braun E (2020) France pushes tougher eu rules for social media in wake of terror attack. *POLITICO*. 26 Oct 2020. <https://www.politico.eu/article/france-renews-social-media-regulation-push-at-eu-level-in-wake-of-terror-attack/>
54. Kearney V (2019) Facebook: dissident republicans Saoradh take legal action. *BBC News*. 30 Jan 2019, sec. Northern Ireland. <https://www.bbc.com/news/uk-northern-ireland-47043375>
55. Letsch C, Rushe D (2014) Turkey blocks YouTube amid 'national security' concerns. 28 Mar 2014, sec. World news. <http://www.theguardian.com/world/2014/mar/27/google-youtube-ban-turkey-erdogan>
56. Loadenthal M (2015) Introduction: like finding a needle in a pile of needles: political violence and the perils of a brave new digital world. *Critical Stud Terrorism* 8(3):456–465. <https://doi.org/10.1080/17539153.2015.1094266>
57. Lonas L (2021) Pompeo, Cruz and other trump allies condemn Twitter's ban on president. *The Hill*. 9 Jan 2021. <https://thehill.com/policy/technology/533486-pompeo-cruz-and-other-trump-allies-condemn-twitters-ban-on-president>
58. Macdonald S, Staniforth A (2021) The tech industry and the regulation of online terrorist content: what do law enforcement think? *Hedayah* (blog). 16 Jan 2021. <https://www.hedayahcenter.org/media-center/latest-news/blog-post-the-tech-industry-and-the-regulation-of-online-terrorist-content-what-do-law-enforcement-think/>
59. Miller S (2010) *The moral foundations of social institutions: a philosophical study*. Cambridge University Press
60. Nicas J, Alba D (2021) Amazon, Apple and Google cut off Parler, an app that drew trump supporters—*The New York Times*. *The New York Times*. 9 Jan 2021. <https://www.nytimes.com/2021/01/09/technology/apple-google-parler.html>
61. Nouri L, Lorenzo-Dus N, Watkin A-L (2021) Impacts of radical right groups' movements across social media platforms—a case study of changes to Britain first's visual strategy in its removal from Facebook to gab. *Stud Confl Terrorism*:1–27. <https://doi.org/10.1080/1057610X.2020.1866737>
62. O'Neill L (2018) Britain first using northern Ireland laws to sue Facebook over censorship claims. *Belfast Telegraph*. 3 Oct 2018. <https://www.belfasttelegraph.co.uk/news/uk/britain-first-using-northern-ireland-laws-to-sue-facebook-over-censorship-claims-37377224.html>
63. Oremus W (2017a) Cloudflare CEO Matthew prince is right: we can't count on him to police internet. *Slate Magazine*. 8 2017. <https://slate.com/technology/2017/08/cloudflare-ceo-matthew-prince-is-right-we-can-t-count-on-him-to-police-online-speech.html>
64. Oremus W (2017b) GoDaddy joins the resistance. *Slate Magazine*. 16 Aug 2017. <https://slate.com/technology/2017/08/the-one-big-problem-with-godaddy-dropping-the-daily-stormer.html>

65. Palzer C (2003) Self-monitoring v. Self-regulation v. Co-regulation. In: Closs W, Nikoltchev S (eds) *Co-regulation of the media in Europe*. European Audiovisual Observatory, pp 29–31
66. Parsons R, Moffat K (2014) Constructing the meaning of social licence. *Soc Epistemol* 28(3–4):340–363. <https://doi.org/10.1080/02691728.2014.922645>
67. Persson T, Roland G, Tabellini G (1997) Separation of powers and political accountability. *Q J Econ* 112(4):1163–1202
68. Phillips R, Edward Freeman R, Wicks AC (2003) What stakeholder theory is not. *Bus Ethics Q* 13(4):479–502. <https://doi.org/10.5840/beq200313434>
69. Prince M (2017) Why we terminated daily stormer. The Cloudflare Blog (blog). 17 Aug 2017. <https://blog.cloudflare.com/why-we-terminated-daily-stormer/>
70. Prince M (2019) Terminating service for 8Chan. The Cloudflare Blog (blog). 5 Aug 2019. <https://blog.cloudflare.com/terminating-service-for-8chan/>
71. Reuters S (2019a) Facebook bans white nationalism, white separatism on its platforms. Reuters. 27 Mar 2019. <https://www.reuters.com/article/us-facebook-hatespeech-idUSKCN1R81ZH>
72. Reuters S (2019b) Italian judge orders Facebook to reopen neo-fascist group’s account. Reuters. 12 Dec 2019. <https://www.reuters.com/article/uk-italy-facebook-neofascists-idUKKBNIYG2AR>
73. Robertson A (2017) Neo-Nazi site moves to dark web after GoDaddy and Google bans. The Verge. 15 Aug 2017. <https://www.theverge.com/2017/8/15/16150668/daily-stormer-alt-right-dark-web-site-godaddy-google-ban>
74. Rodriguez S (2021) Facebook says Donald Trump to remain banned for two years, effective from January 7. CNBC. 4 June 2021, sec. Technology. <https://www.cnbc.com/2021/06/04/facebook-says-donald-trump-to-remain-banned-from-platform-for-2-years-effective-from-jan-7.html>
75. Sadurski W (1999) *Freedom of speech and its limits*. Law and Philosophy Library. Springer Netherlands. <https://doi.org/10.1007/978-94-010-9342-2>
76. Scanlon T (2000) *What we owe to each other*. Belknap Press of Harvard University Press, Cambridge, MA
77. Schauer F (1982) *Free speech: a philosophical enquiry*. Cambridge University Press, Cambridge
78. Schauer F (1985) Slippery slopes. *Harv Law Rev* 99(2):361–383. <https://doi.org/10.2307/1341127>
79. Siapera E, Veikou M (2016) The digital golden dawn: emergence of a nationalist-racist digital mainstream. In: *The digital transformation of the public sphere: conflict, migration, crisis and culture in digital networks*, pp 35–59
80. Statista. n.d. Most used social media 2021. Statista. Accessed 12 June 2021. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
81. Stewart H, Elgot J (2018) May calls on social media giants to do more to tackle terrorism. The Guardian. 24 Jan 2018, sec. Business. <http://www.theguardian.com/business/2018/jan/24/the-resa-may-calls-on-social-media-giants-to-do-more-to-tackle-terrorism>
82. Sunstein CR (1993) *Democracy and the problem of free speech*. Free Press
83. Twitter (2020) The 2020 US elections and Twitter!Twitter Help. <https://help.twitter.com/en/using-twitter/us-elections>
84. Tyko K (2020) Trump Tweet about problems with mail-in ballots quickly labeled by twitter as misleading. USA TODAY. 26 Oct 2020. <https://www.usatoday.com/story/tech/2020/10/26/donald-trump-twitter-mail-ballots-election-tweet-misleading/6049734002/>
85. van Mill D (2017) Freedom of speech. In: Zalta EN (ed) *The stanford encyclopaedia of philosophy*. <https://plato.stanford.edu/archives/win2013/entries/freedom-speech/>
86. van der Vegt I, Gill P, Macdonald S, Kleinberg B (2019) Shedding light on terrorist and extremist content removal. Global Research Network on Terrorism and Technology, paper no. 3. <https://rusi.org/publication/other-publications/shedding-light-terrorist-and-extremist-content-removal>
87. Vibert F (2007) *The rise of the unelected: democracy and the new separation of powers*. Cambridge University Press



88. Vile MJC (2012) *Constitutionalism and the separation of powers*. Liberty Fund, Indianapolis. <https://muse.jhu.edu/book/21621>
89. Vincent J (2019) Facebook Bans UK's biggest far-right organizations, including EDL, BNP, and Britain first. *The Verge*. 18 Apr 2019
90. Wakefield J (2019) Facebook takes down Britain first ads. *BBC News*. 7 Jan 2019, sec. Technology. <https://www.bbc.com/news/technology-46746601>
91. Wellman CH (2009) Rights and state punishment. *J Philos* 106(8):419–439
92. Williams B (1985) Which slopes are slippery. In: Lockwood M (ed) *Moral dilemmas in modern medicine*. Oxford University Press, Oxford, pp 126–137
93. Wong JC (2019) 8chan: the far-right website linked to the rise in hate crimes. *The Guardian*. 5 Aug 2019, sec. Technology. <http://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website>
94. Zuckerberg M (2018) A blueprint for content governance and enforcement. [https://m.facebook.com/nt/screen/?params=%7B%22note\\_id%22%3A751449002072082%7D&path=%2Fnotes%2Fnote%2F&refsrc=http%3A%2F%2Fwww.google.com%2F&\\_rdr](https://m.facebook.com/nt/screen/?params=%7B%22note_id%22%3A751449002072082%7D&path=%2Fnotes%2Fnote%2F&refsrc=http%3A%2F%2Fwww.google.com%2F&_rdr)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# White Knights, Black Armour, Digital Worlds: Exploring the Efficacy of Analysing Online Manifestos of Terrorist Actors in the Counter Terrorism Landscape



Kosta Lucas and Daniel Baldino

**Abstract** Over the past few years, a number of major terrorist attacks have been accompanied by the uploading of detailed, online manifestos, which chart and publicise ideologies, motivations and tactical choices in the backdrop of a dehumanized foe. Such manifestos can also act as inspiration for potential copycats and group-think style supporters within an insulated network. However the types of conclusions that can be drawn from manifesto analysis is a complex issue. The broad aim of this chapter is to explore such identity construction and the usefulness of analysing terrorist manifestos through a narrative framework, with a view to demonstrating that manifestos can be understood as a script to a violent performance (the terrorist act) in the theatre of terrorism (the digital world). These insights can serve the development of policy directed towards aspects of the personal attitudes and the social drivers that are necessary for the amplification of violence rather than in the often impenetrable prediction of who is and who is not likely to become a terrorist actor.

## 1 Introduction

Militant actions and the promotion of online manifestos demonstrates a concerning advancement in efforts to prevent and counter terrorism and violent extremism<sup>1</sup> (“PVE”). Over the past few years, a number of major terrorist attacks have been

---

<sup>1</sup>For the purposes of this chapter, “violent extremism” is defined as the broad umbrella term of ideologically motivated violence, where terrorism is a subset. These two terms may be used interchangeably throughout this paper and will depend on the terminology being used in the respective piece of research literature being quoted.

---

K. Lucas  
Peace and Conflict Studies, University of Sydney, Sydney, Australia

D. Baldino (✉)  
The University of Notre Dame Australia, Fremantle, Australia  
e-mail: [daniel.baldino@nd.edu.au](mailto:daniel.baldino@nd.edu.au)

accompanied by the uploading of detailed, online manifestos, which chart and publicise ideologies, motivations and tactical choices in the backdrop of a dehumanized foe. Such manifestos can also act as inspiration for potential copycats and group-think style supporters within an insulated network. However the types of conclusions that can be drawn from manifesto analysis is a complex issue. From an intervention point of view, the analysis of extremist manifestos can be a fraught exercise, often grappling with a number of issues both theoretically (narrator reliability and analytical subjectivity) and ethically (media reproduction and the magnification of inciteful messaging).

However, recent work in areas like criminology [1, 2] and sociology [3, 4] has suggested that there is merit in exploring identity formation and applying narrative frameworks in the analysis of terrorist manifestos. Manifestos not only provide insights into inhabited social worlds of terrorist actors but can offer a number of benefits in efforts to understand the terrorist mindset and its underpinnings: that is, a sense of how different factors merge to define, shape and sustain activist narratives and biases and then direct identity-formation. As a starting point, individuals who use politically motivated violence often seek to justify it as this level of engagement can feed into a wider sense of identity, meaning and purpose. Identity formation and reification in particular, is an essential aspect of the radicalisation process.<sup>2</sup> In other words, an ideology that espouses or validates violence can give individuals ‘answers’ that make sense for their particular need and create an identity fusion (see [5]).

The broad aim of this chapter is to explore such identity construction and the usefulness of analysing terrorist manifestos through a narrative framework, with a view to demonstrating that manifestos can be understood as a script to a violent performance (the terrorist act) in the theatre of terrorism (the digital world). To this end, the chapter will unpack the dynamic of identity fusion and a specific online terrorist manifesto that coupled with an activist extremist agenda while seeking, in part, to exploit the media in a national security context. The March 15 terrorist attacks in Christchurch in 2019 was staged ‘in real time’ (see [6])—by an adherent to a Far-Right extremist ideology that cut across several transnational movements within the milieu of White supremacy, Neo-Nazism and “ecofascism”<sup>3</sup> (herein the author of this manifesto will be referred to as “BT”).<sup>4</sup>

The chapter will also utilise social identity approaches and the above analytical framework to explore BT’s tendency to search for order, empowerment and structure;

---

<sup>2</sup> In this context, radicalisation itself refers to a process by which an individual or group embraces an extreme ideology so that an ‘outsider group’ is seen as posing a dire threat to the survival of the ‘insider-group’ and they therefore reject the existing status quo – an outlook that might justify the use of violence to bring about political change.

<sup>3</sup> Eco-fascism is an ideology that blames the demise of the environment on overpopulation, immigration, and over-industrialization and has its roots in neo-Nazism as a means to ‘protect and save’ the planet.

<sup>4</sup> This point has been taken to heart by NZ Prime Minister Jacinda Ardern who has not publicly used BT’s name in any communications or statements (<https://www.abc.net.au/news/2019-03-19/christ-church-shootings-jacinda-ardern-house-speech-shooter-name/10917030>).

not only for what their experiences and orientations might reveal about the often-unclear dynamic between online and offline behaviour, but also the cognitive styles of those who inhabit toxic and hateful digital ecosystems that can boost activist identities, negative stereotypes, expressions of ethno-centrism and wider processes of self-justification that might lead to validating the value of violence.

Lastly, a number of other ethical issues arise between media reporting obligations and the content of a terrorist manifesto that is promoting violence. This dilemma will also be briefly explored. In particular, [2] cite the influence that the media has in providing violent actors a platform and the consequences of expanding an extremist profile that can inspire contagion and copycat effects. Media coverage of mass shooters rewards them by making them famous and delivers a clear incentive for future offenders to attack. Instead, the authors argue that if the media modifies how they cover mass shooters, such anticipated changes might be able to deny offenders the personal attention they seek in their quest for significance and help to deter some future perpetrators from normalizing violent behaviour [62].

## 2 Background

The challenges to PVE are represented by the need to recognise and detect routes into violent extremism and indeed proactively work to mitigate processes of radicalisation. This challenge is complicated due to the globalisation of ideas and technological advancements that can link like-minded individuals, promote moral ambiguity and strengthen zero-sum radical beliefs that rationalize the utility of violence. So even with a numerical decrease in the actual occurrences of terrorist violence globally [7] the threat to society by acts of violent extremism consistently ranks as a high level priority for governments all over the world while any ‘silver bullet’ solution towards completely eliminating terrorism is naïve at-best or simply over-simplistic (see [8]). At the same time, there are always trade-offs when considering the search for security—this applies to governance approaches as well as to the legal and ethical aspects of security.

One particular contemporary stream of risk assessment that has re-emerged in recent years as a revised threat to national security and community stability is due to the growth and impact of Right-Wing Extremism (‘RWE’). As captured by the United Nations Counter-Terrorism Committee in April 2020, in exploring unique forms of political violence:

... extreme right-wing terrorist groups and individuals are becoming more transnational. Research has long recognized the potential for extreme right-wing groups to forge strong transnational links and build networks. Recent evidence suggests that there has been a greater exchange of views between like-minded individuals, both online and offline. These connections allow extreme right-wing groups to improve their tactics, develop better counter-intelligence techniques, solidify their violent extremist views and broaden their global networks.

Such extremist movements continue to employ a number of tactics to magnify and amplify messaging, outreach and recruitment that can strengthen extremist identity. And one specific trend that does appear to be on the rise is the online promotion and use of manifestos by lone actors [9]. More than just statements of intent and blatant propaganda, perpetrators of mass violence such as school shooters and violent extremists have often drafted and disseminated manifestos for more personal reasons: i.e., to seek fame or notoriety [2, 10] or to reinvent themselves as white knights in “black armour”. Black Armour is a phrase coined by ([11], 92) is defined as “the process that such individuals may come to embrace a self-styled image based on low self-esteem or negative self-perceptions that may be tinged with an ominous or threatening undertone. That is, they embrace their dark, negative cognitions and fashion them into a recognizable suit of black armour”.

In broad terms, a manifesto is defined as an *ex-ante* communique expressing an actor’s values, intentions and the motivations behind their actions (which at the time of its writing are yet to occur and by the time of dissemination already took place). Rather than incoherent ramblings or crude propaganda, the assumption here is that terrorist manifestos represent an exercise in the selection and emphasis (or de-emphasis) of issues, problems, moral evaluations and solutions, that act to legitimize target selection and a course of violent action (see [12]).

With the increasing ubiquity of the Internet as a way for users to generate and disseminate their own content, the ability for these works to reach wider audiences is unprecedented (see [13]. As summarised by [14], 23) “...this digital ecosystem is fuelling a cumulative momentum which serves to lower ‘thresholds’ to violence for those engaged in this space...as one attack encourages and inspires another, creating a growing ‘canon’ of ‘saints’ and ‘martyrs’ for others to emulate”. So from an audience’s perspective, we assume that a terrorist actor’s intention is to either terrify or inspire us, depending on what type of audience segment we represent to them. Thus in the broad spectrum of policy, research, and programming aimed at preventing such acts from recurring in the future, understanding the terrorist’s mindset, identity formation and interpretation of their social world does remain one of the most central factors for developing appropriate and effective PVE policy responses (see [15], 107).

Much academic interest in the terrorist use of manifestos has tended to tilt towards more psycholinguistic assessments of risks and threats for the purposes of intervention [16]. And while these types of analyses are important (and indeed, foundational) they often face a number of challenges including narrator reliability and analytical subjectivity ([1], 634). These problems can sometimes undermine the usefulness of analysing the narratives for the purpose of risk assessment and intervention, this in addition to the close proximity of the publication to the violent act that leaves little time for preventative actions [17]. In addition to this, and as explored later, violent online actors and actions do pose a difficult challenge for how mass media, social media companies and other public commentators should respond to a massacre broadcast of such manifestos across the Internet.

### 3 Manifestos as the Script, Violence as the Final Act

Few studies have applied narrative frameworks to terrorist manifestos in order to learn about the social worlds that terrorist actors construct within their own interpretations (See [18, 19, 64]). This is despite their potential to teach us about the mindsets, experiences and logic (however, flawed or compromised) that lead to these actors to violence in the first place. Instead, scholars of terrorism have tended to focus on particular aspects such as psychological predispositions or traumatic experiences as necessary precursors to violent radicalisation (see [20]). Yet such research ambitions can fail to view terrorist violence as acts embedded within, and reliant upon, a social context, in-group pressures and a process of moral re-justification to, in part, create an ‘us-against-them’ atmosphere in daily life.

The value then of analysing terrorist manifestos through a narrative framework becomes not just an exercise in understanding what story the author is telling, but also possibly what story the author is living. In other words, terrorist manifestos could also be considered as another way to understand a terrorist’s own narrative of the social world—sequences of linked people, places and processes (see [4])—that could indeed provide and add insights to our ability to answer elusive questions about the motives, processes and synergies that result in a sustained participation in violent extremism. This approach can incorporate a wide range of prevention efforts that aim to curb the potential of generalised imitations and copycat violence ([14], 4).

The assumption to explore is whether a terrorist actor’s social constructs and search for social acceptance become replicated in their manifesto’s rhetorical choices and overall narrative structure that supports violence. If the goal is to simply inspire or terrify, we may assume that a terrorist actor selects and emphasises specific details that would resonate with a prospective like-minded audience. The actor/author is also usually the protagonist, and everyone else—be it a group, an individual, an event of injustice—are invariably represented as antagonists, core plot points and/or catalysing events [21]. The narrative arc is often one of a transformation (of the protagonist) into a warrior for a higher cause and who is no longer paralysed by moral ambiguity. Hence, applying such a rationalisation framework to terrorist manifestos becomes an exercise in exploring the processes that can maintain violent extremism in analysing both the ‘what’ and ‘how’ in that story, including cost–benefit calculations. Such analysis can provide unique insights into how these actors interpret and experience (or at least wish to demonstrate) a sense of control and purpose about themselves, others, and the world around them.

### 4 Cues and Liner Notes: World-Building and Motivations of Terrorist Actors

If there is any degree of consensus in a contentious field of study like terrorism, it is that it will never be possible to fully understand the motivations of terrorist actors, and

why they make a decision to employ violence as a means to an end (see [22], 245). While empirical studies acknowledge this reality, scholars note that there is still a tendency to treat participation in terrorist activity as a syndrome of a fixed etiological state, rather than a decision (however, flawed or compromised the decision-making process is) to utilise the act of terrorism as a tool to achieve specific goals ([23], 193).

Wakefield ([4], 1) notes below that this tendency of academics and policy-makers often manifests as a quest for providing simplified explanations for people's participation in terrorist activities, invariably revolving around one of the following factors:

... brainwashing, cultural factors, frustration aggression, identity crisis, mental illness, narcissism, political exclusion and oppression, rational choice, poverty and relative deprivation all feature prominently within studies of terrorism.

While there are certainly salient themes that recur in case studies of terrorism across disciplines, the reality is that there is no 'one size fits all' explanation that applies to all cases. What applies in one case is not guaranteed to be a factor in another. Hence by pathologizing the actions of terrorist actors as predominately impulsive, or as a manifestation of some evil disease of the mind, we risk overlooking a very important aspect of terrorism: its inherently social nature and a need to have a 'story'—and as explored below with BT, this can incorporate an almost mythological element with their fixation to medieval battles and figures. For instance, BT had a black sun symbol (sonnenrad) emblazoned on their rucksack that might appear innocuous at face-value. But this symbol does hold significant meaning including in white supremacist youth and occultist subcultures (see [24]). Its origins are tied to mosaic floor in a castle where the black-uniformed SS elite would conduct pseudo-religious ceremonies during Nazi Germany.

So as an alternative etiological starting point, sociologists such as ([4], 2) suggest revisiting this issue of violence using a starting point which places a terrorist actor with a distinctive sense of agency—even if this sense of agency is influenced by formative life experiences, and dynamic emotional worlds—and is more constructivist in nature. Much in the same way a narrative is a sequence of plot points linked by a binding thread spanning from its beginning to its end with observable signs of an extremist branding and identity development. Support for this view of terrorism can be seen with some variances from the likes of [18, 25–30].

Smith and Talbot's 'Social Influence Model of Violent Extremism' (SIM-VE) extends the 'people, places and processes' view of terrorism by exploring the ways that social influences and differing processes of radicalization can lead to violent behaviour. As per the SIM-VE model, social influence is conceptualised as the plethora of influences that transform a person's identity to align with a violent extremist group, shape their beliefs to align with an extremist ideology, and reconstruct their moral position to allow for violent action to become ethically acceptable. Security Council Counter Terrorism Committee ([3], 105) identify three broad categories of social influences: ideological (beliefs), behavioural (emotions and cognitions) and social (relational).

Notably these are very broad categories to describe the domains of the behaviour and are not mutually exclusive from one another. In fact, [3] emphasise the rather interdependent nature of each domain, with changes in one area of a person's life contributing to the changes within others. This is particularly important in light of the increased and diffuse connection to toxic online and digital echo-chamber spaces that actors are increasingly showing evidence of inhabiting and developing bonds of friendship within while usually displaying black-and-white perceptions of wider society.

Therefore, the question becomes to what extent a manifesto-embedded narrative bridges the gaps between the people, places and processes of terrorism in any given account. Thematically, the focus of such analysis is to uncover the constructions of self, others, and the world-at-large present within any given manifesto. However to understand how each terrorist actor engages with these themes within their own self-accounts, the analysis not only has to consider the explicit discursive choices made by the authors, but also the narrative techniques employed by each author. This includes the overall narrative sequencing [31], core as well as periphery events and people [21], coherence of characterisations and descriptions [29]; salience and selection of relevant social themes [32]; and paratextual references [21].

In order to analyse the social narratives contained within a terrorist manifesto, we should not only look at 'what story is being told', but also 'how the story is told.' Therefore, the analysis of the empirical data—i.e. the two manifestos chosen for this dissertation—amalgamates a combination of various qualitative methods including thematic, linguistic and narrative analysis. All of these aspects will be utilised to understand how terrorist actors can frame the issues discussed within their respective manifestos.

### Utilising framing theory

According to ([33], 52), in his seminal work on framing theory, *Framing: Toward Clarification of a Fractured Paradigm*:

Framing essentially involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described. Typically frames diagnose, evaluate, and prescribe

As noted, frames are a foundational aspect of narrative-based exercises, particularly as they relate to the self-accounts of terrorist actors. Berntzen and Sandberg [32], Borum [34], Cottee and Hayward [21], Howard et al. [35], Sandberg et al. [29], all use some form of framing theory in their empirical analyses of terrorist accounts, and this chapter framework will continue method.

This is particularly appropriate for understanding the motivations of terrorist actors as terrorist violence relies on the invocation of *collective action frames* ([32], 760). Considered the language of social movements, collective action frames can be found in terrorist narratives because they emphasise all of the important aspects of the terrorist mindset: group identification and the social nature of the problems and

solutions within any given context. Collective action frames, whether explicit (e.g. “we” and “us”) or implied (e.g. putting yourself in the shoes of the characters) what allows a writer of a text to tap into the “emotional raw materials” needed to mobilise people towards a particular goal (see [12], 85).

In order to analyse BT’s manifesto in accordance with a narrative-framing theory approach, the analysis was conducted in the following way. Firstly, the social themes that were identified in the literature review were used as the thematic basis for the analysis of each manifesto. These themes are:

1. Self as the pseudocommando;
2. Intense feelings of anger at injustice and revenge against others;
3. Social experiences of victimisation, isolation and ostracism;
4. Unbalanced existential needs; and
5. The world as “black and white.”

To understand how each manifesto engages with these themes, the analysis not only considered explicit discursive choices made by the authors, but also narrative techniques employed by each author, including: overall narrative sequencing [26], ‘core and periphery’ events and people [21], coherence of characterisations and descriptions [29] and salience and selection of relevant social themes [32].

## 5 Analysing BT’s Manifesto

BT is an Australian man who, at the age of 28, took the lives of 51 Muslim worshippers at the March 2019 Christchurch Mosques’ attacks. The killing spree itself was self-broadcasted to social media, with hundreds of thousands of viewers witnessing the events before the live-stream was taken down approximately hours after its initial broadcast started [11]. Hours prior to committing the terrorist attack, BT had released their manifesto which broadly revolved around the conspiracy theory of ‘The Great Awakening’, whilst drawing on various theme including fear of Muslim conquests (as epitomised in the past by Ottoman rule), White genocide<sup>5</sup> (orchestrated by increasing birth-rates and migration patterns of Muslims and other non-whites into Europe, the US, New Zealand); occurrences of immigrant violence against White Europeans and concerns over overpopulation and eco-fascism<sup>6</sup> all of which fit under the broad umbrella of Far-Right Extremism [19, 38].

BT’s 74-page manifesto can be divided into two distinctive parts: The first section is written in a predominantly Q&A style, addressing hypothetical questions and comments from various imagined or intended audiences. The broad categories of

---

<sup>5</sup> The term “White genocide” refers to a conspiracy theory which alleges a premeditated genocidal campaign against the ‘white race’ by turning it into a minority in its own lands. See: [36].

<sup>6</sup> The term “eco-fascism” refers to the beliefs of “living in the original regions a race is meant to have originated in and shunning multiculturalism is the only way to save the planet they prioritise above all else.” See: [37].



audience can be classified as their “embedded allies”, those who are sympathetic to their cause and hostile to their enemies: “the invaders” (non-White migrants) and “the traitors” (social progressives that are sympathetic towards the so-called “invaders”). The second part of the manifesto could be classed as a series of general calls-to-action, addressing broader ideological issues that seem to underpin the beliefs espoused within, and a sense of interpretation of the reality of, the first part.

Therefore BTs manifesto employed multiple narratives and tropes to justify their attack and to reach multiple audiences to maximize its impact. At the same time, it should be noted that academic research into terrorism and the partial replication of such manifestos in itself is often confronted with cost/benefit breakdowns and beset with related ethical challenges and dilemmas. But in the case of BT, as noted above, we argue that such an analysis of extremist context can identify and extrapolate insights to support different PVE perspectives with content-based analysis predominately intended to help to developing risk profiles as well as prevent future attacks.

### 5.1 *The Self as ‘The Pseudocommando’*

In terms of how terrorist actors construct their notion of self, [11] work on pseudocommando identification that provides a starting point and has been adapted by many prominent researchers such as [2, 39] in subsequent works to understand terrorist behaviour. Based on extensive analyses of perpetrator profiles and self-narratives, Knoll describes pseudocommando identification as a construction of a warrior-like mentality. It is often a self-characterisation that is in equal parts vengeful and narcissistic. Almost paradoxically, said actors will go to great lengths to appear deliberate and rational ([11], 87). Pseudocommando identification can be detected through analysis of both textual and non-textual data and can be expressed in a number of ways:

- Self-references as a soldier, a warrior or other militarised self-characterisations ([16], 249–250),
- A strong identification as an agent for a higher/collective cause or with key figureheads ([16], 249–250),
- A preoccupation with weaponry and war memorabilia ([16], 249–250), and/or
- A painstaking attempt to document the logic and decision-making leading to their decision to use violence—likely an attempt to control the narrative rather than being written off as of unsound mind or merely frustrated/aggrieved ([11], 87).

BT’s manifesto demonstrates a noteworthy pseudocommando self-identification in their methodical detailing of their intentions, their expositional way of addressing hypothetical scenarios and audience questions (exhibiting deliberation and self-awareness) and the flexing of their tactical prowess. While these elements are present throughout the manifesto, BT also writes from the first-person perspective and goes to great lengths to position themselves as an ordinary person (often using the word

“regular” to describe themselves and their circumstances; “an ordinary White man or working class, low income family”) who then transforms into a reluctant hero (“... who decided to take a stand to ensure the future of my people”). Further, BT lists a series of historical events of aggression towards their own people—again, the aggressions of the “invaders” and “traitors” against the “European people” that have been victimised in a so-called clash of civilisations.

Another pseudocommando tendency BT displays is a grandiose self-image, even if reluctantly. BT explicitly claims that they were not motivated by seeking fame, stating it would be “laughable” to do so. Conversely, this attempt to humble himself does ring to some degree hollow when contrasted with the various instances of other fame seeking behaviours and fantasy present elsewhere in the manifesto. They cite affinity with many figureheads of similar movements (Oswald Mosley, founder of the British Union of Fascists and progenitor of modern ultranationalist movements) and other mass shooters and terrorists (Dylann Roof and Anders Breivik). BT even states that their violent actions will be celebrated in the future, citing Nelson Mandela winning the Nobel Peace Prize. This conflicting characterisation and constant re-assessment of morality evokes Lankford’s ([2], 473) observations that the:

evidence [of fame-seeking by mass shooters] requires more interpretation, because even though many have admitted wanting attention and directly orchestrated their attacks to get it, they often claim they want this attention for their cause [and not themselves].

## ***5.2 Intense Anger at Injustice and Seeking Revenge Against Others***

BT went to great lengths to appear deliberate and methodical in the decision to use violence by pointing to a long list of transgressions by respective antagonists within the manifesto. Notably, while BT is dedicated to the idea of ‘tit for tat’ cultural confrontations, their transformation into a paramilitary soldier appears to have resulted from group-think pressures as well as some transformative events that they did witness online, but had not experienced personally. This observation provides an interesting consideration about how zero sum calculations and community (rather than personal) gains can underpin terrorist violence. This is perhaps a deliberate narrative choice by BT, in order to overcome emotional barriers while not appearing impulsive or irrational—an undesirable trait in the pseudocommando mindset [11].

BT did not identify a specific person or group as their antagonist. Instead, the antagonists are defined by generalised albeit hardened stereotypes, negative imagery and the corrosive, corrupt and anti-social values they represented. For BT, it was the “invaders” seeking to replace White Europeans due to below-replacement white birth-rates (paired with the high fertility of non-white immigrants) that will lead to the replacement of the white population in the West. Notably, the manifesto demonstrates a significant highlighting of comradeship and perceived collective or community injustice rather than direct references to personal traumas and experiences of personal

victimisation and isolation. Again, this may confirm Knoll's findings of tendencies for terrorists and mass shooters to try and avoid appearing irrational and illogical. It may also reflect a desire by BT to control their own instinctive narratives and to present himself to an external audience in a distinct way: fearsome to opponents and galvanising to supporters in efforts to inspire other in-group members to action.

Additionally, according to [3] SIM-VE Model, violent behaviour is often conceptualised as an embodied and lived experience with a person's cognitive and sensory dimensions—namely emotions—that influence how a person identifies an enemy and then rationalises and performs violent actions. In regards to the role of emotions and associational drives in fuelling a terrorist mindset [40] notes that anger and the reshaping of identity to create the distance from the other that is, as was noted earlier, necessary to facilitate a cognitive direction towards self-justification, moral authority and violence. Similarly, in a study by ([41], 94), they hypothesise that the interplay of anger (an assessment of another's actions), contempt (an evaluation of another's attributes and worth) and disgust (an evaluation that something or someone is so intolerable, that they must be removed) can provide the emotional powder keg that underlies the acceptance of indiscriminate violent actions.

Interestingly, in line with Knoll's conceptualisation of the pseudocommando as collectors of injustice, BT includes a list of violent actions by the other at several points in the manifesto. While referring to their enemies as "invaders", BT is not simply preoccupied with the demonization of the other through simple name-calling. Instead, they also frame the other as a corrosive yet worthy (and indeed, dominant) adversary that must be countered by any means necessary.

They were an obvious, visible and large group of invaders, from a culture with higher fertility rates, higher social trust and strong, robust traditions that seek to occupy my people's lands and ethnically replace my own people [sic]

Certainly, by positioning the other as possessing a war-like presence, they arguably intensify their own pseudocommando identification—and this psychologically primes themselves to engage in conflict and violence (even against random non-combatants) who they still perceive as guilty of aggressing against white people. While listing certain global events (such as the death of Ebba Akerlund and the 2017 French election) as key impetuses for actions, they do not allude to any specific personal instances of victimisation, humiliation or ostracism. So it is unclear whether BT was subjected to experiences of victimisation and ostracism in their personal life. In the event that they did have these experiences, a deliberate choice appears to have been made to omit such information, again conceivably in order to reinforce themselves as both a fearsome and noble character.

On the first page of BT's manifesto, they do also make it explicitly clear that the reasons for their terrorist attacks were done in the name of vengeance: against the "invaders"; against "Islamic slavers", as payback for "enslavement" and "murder" of their people on Western lands. In the Q&A segment of their manifesto, BT addresses the hypothetical accusation that they are "a bigot, racist, xenophobe, islamphobe, nazi, fascist [sic]". Their response, at first instance, appears to be a vitriolic tirade, ultimately ending with "you're fucking dead, kiddo."—a direct quote directed at the

reader who is assumed to be an enemy or antagonist. BT's response to this question is in fact a well-known quote that was written to satirise the vitriolic online behaviour of people within the gaming community.<sup>7</sup> Further, while they tend to utilise the laundry list method of building to build their case for violence against the other, and a patriotic tone when addressing hypothetical supporters, BT appears to stoke their own rage when addressing those perceived as traitors including other white people who have essentially turned on their own culture and allowed a white genocide to occur.

The only other distinctive outpouring of emotion that rivals this above passage was BT's own self-described turning point when they visited the graves of fallen soldiers during travels through France. After highlighting the impact of French elections in 2018, BT demonstrates an utter loss of faith in the establishment of a "once great" European nation, which essentially betrays its citizens:

The candidates were an obvious sign of our times: a globalist, capitalist, egalitarian, an ex-investment banker has no national beliefs other than the pursuit of profit versus a milquetoast, feckless, civic nationalist, an uncontroversial figure who's most brave and inspired idea resolved to the possible deportation of illegal immigrants

Given their frequent references to the opportunism of the "invaders" and lamentation of White Europeans complacency, these passages may possibly be most revealing of one of the main sources of rage, and perhaps, a strong sense of empowerment in that face of an us-against-them dynamic.

In terms of their preoccupation with revenge, BT also employs gendered characterisations when highlighting inspirations for violence, and in particular, the idea of women as damsels in distress (as in the case of Ebba Åkerlund, the 11-year-old girl killed in the 2017 Stockholm terror attack) and as muses (mentioning African American conservative commentator, Candace Owens whom they cite as having had the strongest impact on their radicalisation). Generally, manifestos that use gendered characterisations may provide an insight into the identity construction of terrorist actors, particularly in the case of ideologies that have very narrow confines for hegemonic masculinity—namely, entitlement and expectation [42, 72]. In this case, while BT's construction of women is consistent with more traditional, passive Right-Wing extremist constructions, it could be a revealing window into their own personal needs, as they frequently use female characters as sources of inspiration and encouragement.

Interestingly, BT did not engage with any themes relating to social victimisation, isolation or ostracism in their own life. And this may be for a number of reasons. A likely explanation would have been the fact that including any details of this nature could undermine self-directed grandiose characterisations as a "pseudocommando". Thus inclusion of such personal, painful, and emotional experiences would contradict the pseudocommando framing as methodical, deliberate and unemotional. It also may have interfered with the resonance of any collective call-to-action framing, if the grievance appeared too personal to a prospective reader (see [32]). Of course,

---

<sup>7</sup> See: Know Your Meme. 'Navy Seal Copypasta | Know Your Meme'. Accessed 5 October 2019. <https://knowyourmeme.com/memes/navy-seal-copypasta>.

it could also be that BT may not have been consciously aware of the psychological effect that such experiences would have had on them.

But overall BT did go to great lengths to position themselves as an average bystander that was transformed into a warrior by the changing world around them. This was evident in their self-framing as coming from an unremarkable upbringing, through to eventually justifying their acts as objective, reasoned choices, which were forced on them by confronting circumstances. Naturally, this depiction explicitly accords with the pseudocommando tendency to consider oneself as ‘collectors of injustices’. This, contrasted with the absence of any mention of personal experiences of victimisation, ostracism and isolation, leads us to interrogate a terrorist actor’s attachment to the issue of injustice, grievance and feelings of efficacy to begin with.

### 5.3 *Unbalanced Existential Concerns*

Another notable feature is the fact that the manifesto demonstrates a worldview which sees the world as place where violence is not only constant, but inevitable. BT was explicit about a gloomy ‘only the strong survive’ outlook, with all outgroup actions being framed (and usually misapplied) as war-like and aggressive in nature.

In BT’s manifesto, their dehumanising perspectives and framing of anger at the injustices inflicted towards white people is intertwined with their framing of broader existential concerns regarding purpose (evolving from “why won’t someone do something?” to “why don’t I do something?”) as well as the injustices they bore witness to on their travels—which resembled a quest for meaning, even if the search for cognitive justification is not directly expressed that way.

To date, much research has focussed on more deficit-oriented risk factors such as negative emotions and experiences; however, some scholars have noted the need to consider the influence of more propulsive emotional drivers that we would normally consider as “positive.” For example, ([43], 965) note that certain feelings experienced by terrorist actors are overlooked when trying to understand a decision-making process. Arguably, a more philosophical aspect of the extremist mindset and emotional state, existential concerns refer to the fundamental questions of existence itself that may be motivating factor in an actor committing an act of violence in the name of a cause or ideology:

The key argument advanced in what follows is that terrorism, for those who practice and embrace it, can be profoundly thrilling, empowering and spiritually intoxicating, and that this particular aspect of it may inform, along with other key motivations no doubt, the decision to engage in it. (p. 965)

Crenshaw [43] highlight the tendency of the literature to pinpoint deficit-based vulnerabilities and a loss of control over one’s life as leading causes as driving an actor down the road towards anti-social behaviour and violence. In their view, dehumanising perceptions of others is often performed at the expense of moral clarity

while processes of identification can be driven by the attainment of general feelings of empowerment and recognition.

These too, are basic and possibly more propulsive emotional states, than anger at injustice alone. Others, such as [44] adapt the work of existential psychologist, Irvin D Yalom, and put forward a framework of understanding of the terrorist mindset as a way to mitigate against the core existential concerns of human existence: *identity, isolation, death, freedom and meaning*. McBride ([41], 561) also makes a philosophical case for understanding the logic of involvement in terrorism, but draws more explicitly on terror management theory and existential psychology as a quest to alleviate existential anxiety:

... terrorism may be driven by an existential-terroristic feedback loop: a cycle in which people support or engage in terrorism to alleviate existential anxiety but ultimately find this anxiety exacerbated in the wake of the violence they create or sanction. The loop is closed when this exacerbated anxiety compels them to reaffirm their support of, or participation in, terrorist violence.

Such existential concerns do add a potentially metaphysical aspect to understanding the motivations of terrorist actors because they deal with the fundamental orientations of one's world views, and the emotions and events that influence someone's path towards (or away) from death essentially. Given that the mindset of terrorist actors is generally thought of as rigid and 'all-or-nothing'-oriented, these findings do seek to encourage future researchers to pay more attention to the existential concerns of terrorist actors, as the scope of their motivations may be far more intangible than they might sometimes superficially appear.

Overall, BT's manifesto details a long search for existential meaning and purpose. Their sense of place, while not especially a focal point, is conspicuously framed as a reflection of the society/culture/order of the race they believe they "belong" to. In terms of Australia and NZ, BT sees them as both extensions of Europe and as the "last existing Utopias for White Europeans". And while they acknowledge that they have no issues with Muslims who are practicing their faiths in their own lands, they are reliant on dehumanizing perceptions and quick to attribute the degenerate, unappealing nature of the enemy.

It is worth noting that BT indicated that the online documents that they authored might be a last chance at to fulfil previously unfulfilled aspirations. Certainly, BT demonstrated an explicit preoccupation with the military, constructing their manifesto as a tactical document and exposing the reader to hypothetical operational situations. This may have been another way to inform readers that that they had thought of all possible scenarios and that the eventual terrorist action approach was indeed the most effective and judicious.

And these existential concerns do dovetail into the dogmatic and even self-protective cognitive aspects of the terrorist mindset. For instance, scholars like ([45], 32) and ([31], 205) assert that a rigid, fixed and unbending cognition often forms to provide the backbone for the necessary commitment to violent action in a violent extremist context.

In other words, a tendency to want to pathologize terrorist actors as 'sad, mad or bad' often betrays the role of identity development and the complex social reality of

factors that sustain involvement in terrorism. Researchers can emphasise the need to understand how individuals use their idiosyncratic social experiences to construct their worlds with, and how and why someone could decide upon the need to commit acts of violence in the name of an extremist ideology. Such a point of view is particularly confronting for policymakers, as the natural implication of this finding is to try and look at such morally reprehensible actions with a form of cleared-eyed rationality.

In sum, BT's manifesto possesses a strong dogmatic intolerance and tendency to dehumanize those who opposed their beliefs. They frequently employ clear-cut binaries on the core issues like nationalism and hierarchy. For example, some clear examples related to White genocide include: the strong association made between "birth-rates", "fertility rates" and the presence of "invaders"; coupling the victimisation and "rape of White women" with expressions of vengeance and expressions of rage against those they consider "race traitors." BT views the world as an inherently violent place, full of prejudice and perpetually at war and their political standpoint is a desire for group-based dominance along a superiority-inferiority racial and gender dimension. Critically, a pseudocommando identity and a need for cognitive closure provided a lens to feed into an all-encompassing sense of purpose, belonging and moral disengagement.

## **6 The (Digital) World We Live in or the (Digital) Battleground We Fight in?**

In regards to the transmission of extremist ideas, when we think of the Internet in relation to terrorism, we can accept that the digital era has opened up many routes to spread hateful ideology from subcultures to general audiences as well as create spaces for like-minded individuals to come together and form a sense of community that they do not have access to in their offline circumstances.

The RAND Corporation [46] findings about the role of the Internet and violent extremism, provides a useful starting point for interrogating the assumptions we make around the relationship between a terrorist actor's online and offline behaviours. That is, that the Internet:

- did create more opportunities to be radicalised;
- did act as an echo chamber;
- did not necessarily accelerate the process of radicalisation;
- did not replace the need for physical contact in the radicalisation process; and
- did not increase the opportunity for self-radicalisation.

Outside the manifesto itself, BT's extended connection to online extremist groups and forums has been well documented [19]. And as mentioned, the terrorist attack perpetrated by BT was live-streamed via Facebook with the manifesto disseminated hours prior to the event. Yet in terms of BT's manifesto itself, the narrative paints

a highly intermingled and multifarious relationship between BT and their online environment—one that is at once both intentional and filled with propaganda but also coalesced with processes of identification and the search for meaning and direction. Certainly, BT appeared to have been well acquainted with extremist corners of the Internet that repeated, for example, false propaganda about immigrants as “invaders”. In their own words:

From where did you receive/research/develop your beliefs?

The internet, of course. You will not find the truth anywhere else.

While not the prime focus of the manifesto, BT did construct their relationship to the Internet as frequently celebrating it as a liberator from the blinkered, limited preferences that shaped their actions and moral standards in offline circumstances—referring to it as responsible for “breaking the grip of the media on the zeitgeist of modernity”; in short a place to find the ‘truth’. Additionally, BT did appear to be acutely attuned to the online instincts and internet culture of their intended audience, with numerous references to popular memes and frequently utilising in-group vernacular and LEET<sup>8</sup> speak throughout the manifesto. Many of the memes inserted into the manifesto did aim to bolster the fervour and enthusiasm of fellow white supremacists.

Unquestionably, the Internet’s precise role in the process of radicalisation is vexing. But radicalisation does remain a deeply social process. So, in general terms, the Internet allows alienated and disaffected people to find and connect with each other. It also provides a space for those looking for acceptance, recognition and a sense of approval. And some of the most extreme forms of dialogue, including dehumanising and hateful ideas that target specific pre-existing biases, this pattern can become self-reinforcing. In that regard, it can be argued the Internet should remain to be seen as a mechanism to enable or facilitate radicalisation. But as highlighted in the study of the above manifesto, such processes of radicalisation do remain complex and contested. It will incorporate a combination of online and offline communication and a fluid mix of different political, psychological and social factors.

At the same time, media and related reporting frameworks must carefully consider the history of mass shooters seeking fame for their actions and reflect on possible changes to media coverage in the future. Covering mass shootings is a tricky proposition for the media. Indeed, there are entire RWE communities dedicated to spreading propaganda and misinformation. Lauland et al. [47] identify a number of consequences of media coverage, including perpetrators’ fulfilment and incentive to achieve notoriety, competition among offenders to maximize victim fatalities and copycat and contagion impacts in various types of aggressive behaviour stemming from impressionable individuals. The authors cite information from the 2007 Nebraska mall shooter, the 2011 Tuscon shooter, the Virginia Tech shooter, and the Columbine shooters that all suggest that fame and notoriety were a large factor in their decision to engage in mass violence. As a consequence, some such as [48] propose that media refrain from using names and images of current and past shooters

---

<sup>8</sup> LEET is a style of typing that replaces English letters with similar-looking numbers or symbols.



while reporting all other aspects of the story in as much detail as possible. Additionally, as captured by ([12], 13), "... a broader discussion is needed on the merits and drawbacks of internet censorship, particularly regarding the sites and servers which have typically hosted these manifestos". With the modern-day growth of transnational far-right movements, as a starting point, the media must consider the merits of publishing any 'call-to-arms' information that might empower future terrorists.

So while research does suggest that terrorists are often motivated by the notoriety they gain in the media, and manifestos such as BT's—immediately banned in New Zealand and more recently targeted for takedowns by social media companies—can spawn copycats, a number of difficult ethical and associated reporting issues do still simultaneously remain.<sup>9</sup> For instance, some have questioned the ban, stating that it risks turning BT into a martyr and therefore lending a form of legitimacy to their far-right ideology. Stephen Franks argued that the "... damage and risks are greater from suppressing these things than they are from trusting people to form their own conclusions and to see evil or madness for what it is" (cited in [9]). Such questions surrounding censorship and how traditional and social media companies respond to such events will remain provocative—at the very least, in tackling what can be done to minimize the destructive appeal of extremist ideas without infringing on freedom of speech and without limiting the audience's rights to be informed. As captured by Dr Bharath Ganesh (cited in [12]), a researcher at the Oxford Internet Institute in questioning the public purpose of journalism:

Taking down the video [of the Christchurch shooting] is obviously the right thing to do, but social media sites have allowed far-right organisations a place for discussion and there has been no consistent or integrated approach to dealing with it. There has been a tendency to err on the side of freedom of speech, even when it is obvious that some people are spreading toxic and violent ideologies.

Similarly, the UK's head of counter-terrorism policing Neil Basu previously challenged the media to have a "sensible conversation" how they report terrorism, stating that the problem extends beyond social media platforms where perpetrators can disseminate manifestos and convey their crimes in real time (see [15]). Basu cited a report, *Terrorism and the Mass Media*, which had again underlined that there was a need for more lucid and consistent ethical guidelines to help the media in reporting terrorist attacks, in this case similar to the code of practices often used in the treatment of suicide cases. "The key is to find a balanced approach that reduces negative impact, increases positive impact, and enshrines media independence and the public's right to know" (cited in [49]).

Consequently, while a core PVE challenge is not so much countering the active use of violent social media but in inhibiting and averting the conditions conducive to such violence before it happens, there is at least an acknowledgement now that multi-stakeholder, multifaceted approaches are needed in order to respond effectively to problematic content and such debate will entail, at the very least, how to make ethical decisions about what to censor, de-platform, demonetise, leave up, moderate or refer to the authorities. It is worth noting that the Christchurch Call to Action

---

<sup>9</sup> In March 2019, New Zealand banned the possession or sharing of the Manifesto.

and the inception of the Global Internet Forum to Counter Terrorism (GIFCT) had both pushed for systemic changes in the covering extremist content after the attack by reinforcing ‘better practices’, that encompassed not naming the perpetrator/s as well as the reviewing the effects and impacts of commercially sensitive algorithms, to reduce future manipulation and harm. Problematically, despite encouraging wider cooperation between the tech sector and governments in responding to terrorist incidents, the US, Russia and China all remained notably absent in supporting the above initiative (see [50]). Nonetheless, at least in the US case, when a mob of pro-Trump rioters stormed the US Capitol on January 6 2021 and footage of the insurrection flooded social media, some domestic debate was again resurrected about social media as a extremist ‘rallying cry’ and related issues about fact-checking, digital citizenship, media literacy and how tech companies might take a more effective and ethical approach to scrutinising their platforms (see [46]).

## 7 Conclusion

The narrative accounts of extremist manifestos not only provide insights into the social worlds of terrorist actors but they can also offer clear benefits to more sophisticated understanding of the terrorist mindset—including how a complex range of social, political and other related factors might all come together to shape and amplify the terrorist worldview. In short, a core challenge “... is to ask why, when people tell their story, they use or repeat particular phrases to the exclusion of all other possibilities” ([51], p.144). If a terrorist actor’s cognition is an internalised story that is influenced by experiences and emotions resulting from both a mix of online and offline social interactions, then manifestos arguably re-externalise a particular cognitive framework—and can provide policymakers and PVE experts with clues for better targeted prevention and intervention policy approaches.

By distinguishing the central themes in the manifestos of known-terrorist actors, models of radicalisation can begin to create a preliminary understanding of the construction of an actor’s social world that sustains their violent extremism. For instance, we may assume that terrorist actors tend towards possessing some combination of the following experiences:

- high expectations of their own value and worth, and a low estimation of the people they consider themselves different to;
- grievances with their current circumstances and processes that lead them to their current place (usually interpreted through the lens of injustice);
- have social experiences of victimisation, ostracism and isolation that reinforce their internal sense of alienation to other people and places;
- rigid ideas of how the world operates, their place within it and what needs to be done in order to improve their place in it;

- reliance on the Internet as both a tool for promoting terrorist violence and for meeting their unmet social needs of their offline lives, characterised as banal, yet silently tumultuous.

Viewed separately, the above reflections could be considered deeply human and even relatable experiences that are highly applicable to a variety of people in a particular context. However, taken together, they can form a militant identity, a violent lifestyle and a more dogmatic interpretation of the world that is characterised by an extreme social and cultural polarisation. One in which BT's evaluations of people (themselves and Others), places (the world) and processes (their indoctrination) tend to be dichotomous and options for recourse to 'shift' the status quo are perceived as limited and will necessitate and rationalize violence.

Thus analysing terrorist manifestos can serve the development of policy directed towards aspects of the personal attitudes and the social drivers that are necessary for the amplification of violence rather than in the often impenetrable prediction of who is and who is not likely to become a terrorist actor. The narrative accounts of terrorist actors do offer valuable insights into identity fusion and the 'how' and 'why' of their beliefs, associated risks and decisions to engage in violent tactics. Manifestos are not an account of absolute truth and are certainly a form of propaganda. Yet even deliberate selections and omissions can be extremely illuminating. As a result, manifestos should be seen as an interpretation of the experiences that help to shape the transformational opportunities and the related decisions of at-risk individuals as well as those that engage in terrorist violence. The careful study of online manifestos can provide a fertile pathway provide a more holistic account of terrorism in a social media age. This includes the identification of 'red flags' to help to guide the understanding of PVE efforts that will incorporate the demand side of radicalization and the wider environmental and social contexts that can make individuals more receptive to extreme ideas.

## References

1. Harley N (2020) UK's head of counter-terrorism says publicity raises fear of copycat killers. The National News, May 12, 2020. <https://www.thenationalnews.com/world/uk-s-head-of-counter-terrorism-says-publicity-raises-fear-of-copycat-killers-1.1018086>
2. Lankford A (2018) Identifying potential mass shooters and suicide terrorists with warning signs of suicide, perceived victimization, and desires for attention or fame. *J Personality Assess* 100(5): 471–482. <https://doi.org/10.1080/00223891.2018.1436063>
3. Security Council Counter Terrorism Committee (CTC) (2020) CTED trends alerts, United Nations, (April). [https://www.un.org/sc/ctc/wp-content/uploads/2020/04/CTED\\_Trends\\_Alert\\_Extreme\\_Right-Wing\\_Terrorism.pdf](https://www.un.org/sc/ctc/wp-content/uploads/2020/04/CTED_Trends_Alert_Extreme_Right-Wing_Terrorism.pdf)
4. Wakefield J (2019) Christchurch shootings: social media races to stop attack footage. BBC News, 16 March, 2019. <https://www.bbc.com/news/technology-47583393>
5. Halperin E (2015) Changing feelings to promote peace: emotion regulation as a new path to conflict resolution. In: *Emotions in conflict: inhibitors and facilitators of peace making*, Taylor and Francis, pp 155–182

6. Coaston J (2019) The New Zealand shooter's manifesto shows how white nationalist rhetoric spreads. *Vox*, March 18, 2019. <https://www.vox.com/identities/2019/3/15/18267163/new-zealand-shooting-christchurch-white-nationalism-racism-language>
7. Institute for Economics and Peace (2019) Global terrorism index 2018. Institute for Economics and Peace, Sydney, NSW
8. Macklin G (2019) The Christchurch attacks: livestream terror in the viral video age. *CTC Sentinel* 12(6):23
9. OZY.1808 (2019) Special briefing: rise of the terror manifesto. OZY (blog), August 5, 2019. <http://www.ozy.com/need-to-know/special-briefing-rise-of-the-terror-manifesto/95980>
10. Bushman BJ (2018) Narcissism, fame seeking, and mass shootings. *Am Behav Scientist* 62(2):229–241. <https://doi.org/10.1177/0002764217739660>
11. Knoll JL (2010a) The “pseudocommando” mass murderer: part I, the psychology of revenge and obliteration. *J Am Acad Psychiatry Law Online* 38(1):87–94
12. Wright-Neville D, Smith D (2009) Political rage: terrorism and the politics of emotion. *Global Change Peace Secur* 21(1):85–98
13. Copeland S (2018) Telling stories of terrorism: a framework for applying narrative approaches to the study of militant's self-accounts. *Behav Sci Terrorism Political Aggression* 1–22. <https://doi.org/10.1080/19434472.2018.1525417>
14. Manavis S 2018 Eco-fascism: the ideology marrying environmentalism and white supremacy thriving online. *New Statesman*, September 21, 2018. <https://www.newstatesman.com/science-tech/social-media/2018/09/eco-fascism-ideology-marrying-environmentalism-and-white-supremacy>
15. Smith D, Talbot S (2019) How to make enemies and influence people: a social influence model of violent extremism (SIM-VE). *J Policing Intell Counter Terrorism* 14(2):99–114. <https://doi.org/10.1080/18335330.2019.1575973>
16. Colborne M (2019) The Christchurch shooting, Eastern Europe's far-right and a cherry-picked reading of history. *ABC News*, April 10, 2019. <https://www.abc.net.au/news/2019-04-10/christchurch-shooting-far-right-groups-in-ukraine-eastern-europe/10983542>
17. Archer RP (2018) Limitations in the prediction of mass violence: cautionary tales. *J Personality Assess* 100(5):451–458. <https://doi.org/10.1080/00223891.2018.1477786>
18. Ferguson N, McAuley JW (2020) Staying engaged in terrorism: narrative accounts of sustaining participation in violent extremism. *Front Psychol* 11:1338
19. Sandberg S (2013) Are self-narratives strategic or determined, unified or fragmented? Reading Breivik's manifesto in light of narrative criminology. *Acta Sociologica* 56(1): 69–83. <https://doi.org/10.1177/0001699312466179>
20. Conway M (2017) Determining the role of the internet in violent extremism and terrorism: six suggestions for progressing research. *Stud Conflict Terrorism* 40(1):77–98. <https://doi.org/10.1080/1057610X.2016.1157408>
21. Cottee S, Hayward K (2011) Terrorist (E) motives: the existential attractions of terrorism. *Stud Conflict Terrorism* 34(12):963–986. <https://doi.org/10.1080/1057610X.2011.621116>
22. Jenkins BM (1980) The terrorist mindset and terrorist decision-making: two areas of ignorance. *Terrorism* 3(3–4):245–250. <https://doi.org/10.1080/10576108008435463>
23. Kruglanski AW, Fernandez JR, Factor AR, Szumowska E (2019) Cognitive mechanisms in violent extremism. *Cognition* 188:116–123. <https://doi.org/10.1016/j.cognition.2018.11.008>
24. Corner E, Bouhana N, Gill P (2019) The multifinality of vulnerability indicators in lone-actor terrorism. *Psychol Crime Law* 25(2):111–132. <https://doi.org/10.1080/1068316X.2018.1503664>
25. Allely CS, Faccini L (2017) “Path to intended violence” model to understand mass violence in the case of Elliot Rodger. *Aggression Violent Behav* 37:201–209. <https://doi.org/10.1016/j.avb.2017.09.005>
26. da Silva R, Fernández-Navarro P, Gonçalves MM, Rosa C, Silva J (2019) Tracking narrative change in the context of extremism and terrorism: adapting the innovative moments coding system. *Aggress Violent Behav* 47:204–214. <https://doi.org/10.1016/j.avb.2019.05.002>

27. Klonick K (2019) Inside the team at Facebook that dealt with the Christchurch shooting. 25 April, 2019. <https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting>
28. Kruglanski AW, Fishman S (2006) The psychology of terrorism: “syndrome” versus “tool” perspectives. *Terrorism Political Violence* 18(2):193–215. <https://doi.org/10.1080/09546550600570119>
29. Sandberg S, Oksanen A, Berntzen LE, Kiilakoski T (2014). Stories in action: the cultural influences of school shootings on the terrorist attacks in Norway. *Crit Stud Terrorism* 7(2):277–296. <https://doi.org/10.1080/17539153.2014.906984>
30. Vergani M, Iqbal M, Ilbahar E, Barton G (2018) The three ps of radicalization: push, pull and personal. A systematic scoping review of the scientific evidence about radicalization into violent extremism. *Studies Conflict Terrorism* 1–32. <https://doi.org/10.1080/1057610X.2018.1505686>
31. Dean G (2014) Neurocognitive risk assessment for the early detection of violent extremists. Springer Briefs in Policing, Springer International Publishing. <https://www.springer.com/gp/book/9783319067186>.
32. Berntzen LE, Sandberg S (2014) The collective nature of lone wolf terrorism: Anders Behring Breivik and the anti-islamic social movement. *Terrorism Political Violence* 26(5):759–779. <https://doi.org/10.1080/09546553.2013.767245>
33. Entman RM (1993) Framing: toward clarification of a fractured paradigm. *J Comm* 43(4):51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
34. Borum R (2003) Understanding the terrorist mind-set. *FBI Law Enforcement Bull Perspect Am Psychol Assoc*. <https://doi.org/10.1037/e318402004-002>
35. Howard T, Poston B, Benning SD (2019) The neurocognitive process of digital radicalization: a theoretical model and analytical framework. *J Deradicalization* 19:25
36. Perry N (2019) Banning of shooter’s manifesto raises free speech debate in New Zealand. *The Slat Lake Tribute*, March 24, 2019
37. Martin L, Smee B (2019) What do we know about the Christchurch attack suspect? *The Guardian*, March, 15, 2019. <https://www.theguardian.com/world/2019/mar/15/rightwing-extremist-wrote-manifesto-before-livestreaming-christchurch-shooting>
38. Matsumoto D, Hwang HC, Frank MG (2017) Emotion and aggressive intergroup cognitions: the ANCODI hypothesis’. *Aggressive Behav* 43:93–107
39. Cohen K, Johansson F, Kaati L, Mork JC (2014) Detecting linguistic markers for radical violence in social media. *Terrorism Political Violence* 26(1):246–256. <https://doi.org/10.1080/09546553.2014.849948>
40. Langman P (2018) Different types of role model influence and fame seeking among mass killers and copycat offenders. *Am Behav Sci* 62(2):210–228. <https://doi.org/10.1177/0002764217739663>
41. McBride MK (2011) The logic of terrorism: existential anxiety, the search for meaning, and terrorist ideologies. *Terrorism Political Violence* 23(4):560–581. <https://doi.org/10.1080/09546553.2011.575486>
42. Ware J (2020) Testament to murder: the violent far-right’s increasing use of terrorist manifestos. *ICCT Policy Brief*, March, 2020. <https://icct.nl/app/uploads/2020/03/Jaocb-Ware-Terrorist-Manifestos2.pdf>
43. Crenshaw M (2000) The psychology of terrorism: an agenda for the 21st century. *Polit Psychol* 21(2):405–420
44. Pfeifer B, Ganzevoort R (2017) Tell me why? Existential concerns of school shooters. *Religious Educ* 112(2):123–35. <https://doi.org/10.1080/00344087.2016.1113908>
45. Department of Prime Minister and Cabinet (2010) Counter terrorism white chapter: securing Australia—protecting our borders. White chapter. Australian Government, Canberra, Australia. [https://www.dst.defence.gov.au/sites/default/files/basic\\_pages/documents/counter-terrorism-white-chapter.pdf](https://www.dst.defence.gov.au/sites/default/files/basic_pages/documents/counter-terrorism-white-chapter.pdf)
46. RAND Corporation (2013) Radicalisation in the digital era: the use of the internet in 15 cases of terrorism and extremism. RAND Corporation, Santa Monica, CA

47. Lauand A, Moroney J, Rivers J, Bellasio J, Cameron K (2019) Countering violent extremism in Australia and Abroad: a framework for characterising CVE programs in Australia, the United States, and Europe. RAND Corporation, Santa Monica, CA. <https://doi.org/10.7249/RR2168>
48. Moses AD (2019) 'White genocide' and the ethics of public analysis'. *J Genocide Res* 21(2):201–213. <https://doi.org/10.1080/14623528.2019.1599493>
49. Holbrook D, Taylor M (2017) Terrorism as process narratives: a study of pre-arrest media usage and the emergence of pathways to engagement. *Terrorism Political Violence* 1–20. <https://doi.org/10.1080/09546553.2017.1341879>
50. Burton J (2020) Christchurch's legacy of fighting violent extremism online must go further—deep into the dark web. Conversation, March 20, 2020. <https://theconversation.com/christchurch-legacy-of-fighting-violent-extremism-online-must-go-further-deep-into-the-dark-web-133159>
51. Halsey M (2017) Narrative criminology. In: Deckert A, Sarre R (eds) *The Palgrave handbook of Australian and New Zealand criminology, crime and justice*, Springer International Publishing, Cham, Switzerland, pp 633–647. [https://doi.org/10.1007/978-3-319-55747-2\\_42](https://doi.org/10.1007/978-3-319-55747-2_42)
52. Satterfield S (2016) Livy and the pax deum. *Classical Philol* 111(2):165–176
53. Aly A, Taylor E, Karnovsky S (2014) Moral disengagement and building resilience to violent extremism: an education intervention. *Stud Conflict Terrorism* 37(4):369–385. <https://doi.org/10.1080/1057610X.2014.879379>
54. Australian Security Intelligence Organisation (2018) ASIO annual report 2018–2019. Australian Government, Canberra, Australia. <https://www.asio.gov.au/sites/default/files/2018-19%20Annual%20Report%20WEB.pdf>
55. Becker MH (2019) When extremists become violent: examining the association between social control, social learning, and engagement in violent extremism. *Stud Conflict Terrorism* 1–21. <https://doi.org/10.1080/1057610X.2019.1626093>
56. Bilardi J (2015) From Melbourne to Ramadi: my journey. From the Eyes of a Muhajir (blog), 13 January. <https://quadrant.org.au/opinion/qed/2015/06/jake-bilardis-deleted-blog/>
57. Freestone M (2017) Personality, identity, risk and radicalisation. *Int Rev Psychiatry* 29(4):310–312
58. Knoll JL (2010b) The “pseudocommando” mass murderer: part II, the language of revenge. *J Am Acad Psychiatry Law Online* 38(2):263–272
59. Know Your Meme (2019) Navy seal copy-pasta I know your meme. Accessed 5 November 2019. <https://knowyourmeme.com/memes/navy-seal-copy-pasta>
60. Kubiak A (2004) Spelling it out: narrative typologies of terror. *Stud Novel* 36(3):294–301
61. Lankford A (2016) Fame-seeking rampage shooters: initial findings and empirical predictions. *Aggress Violent Behav* 27:122–129. <https://doi.org/10.1016/j.avb.2016.02.002>
62. Lankford A, Madfis E (2018) Don't name them, don't show them, but report everything else: a pragmatic proposal for denying mass killers the attention they seek and deterring future offenders. *Am Behav Sci* 62(2):260–279. <https://doi.org/10.1177/0002764217730854>
63. Meindl JN, Ivy JW (2017) Mass shootings: the role of the media in promoting generalized imitation. *Am J Public Health* 107(3):368–370. <https://doi.org/10.2105/AJPH.2016.303611>
64. Mykietiak C (2016) Fragile masculinity: social inequalities in the narrative frame and discursive construction of a mass shooter's autobiography/manifesto. *Contemp Soc Sci* 11 (4):289–303
65. Pfundmair M (2019) Ostracism promotes a terroristic mindset. *Behav Sci Terrorism Political Aggression* 11(2):134–148. <https://doi.org/10.1080/19434472.2018.1443965>
66. Purtill J (2021) Weeks of rhetoric online made the January 6 storming of the Capitol 'entirely predictable', experts say. ABC News, January 8, 2021. <https://www.abc.net.au/news/science/2021-01-08/us-capitol-storming-planned-for-weeks-online-social-media/13039302>
67. Reis S, Martin B (2018) Psychological dynamics of outrage against injustice. *Can J Peace Conflict Stud Peace Res* 40(1):5–23
68. Rundle G (2019) Not lone wolf nor automaton: the Christchurch terrorist doesn't fit old narratives'. Crikey (blog), March, 18, 2019. <https://www.crikey.com.au/2019/03/18/brenton-tarrant-christchurch-narrative/>

69. Shaw D (2019) Christchurch shootings: UK media must deny terrorists a voice. BBC News, March 20 2019. <https://www.bbc.com/news/uk-47645044>
70. Varvin S (2017) Fundamentalist mindset. *Scand Psychoanalytic Rev* 40(2):94–104. <https://doi.org/10.1080/01062301.2017.1386010>
71. Vertigans S (2013) *The sociology of terrorism: people*. Routledge, Places and Processes. <https://doi.org/10.4324/9780203855812>
72. Vito C, Admire A, Hughes E (2018) Masculinity, aggrieved entitlement, and violence: considering the Isla vista mass shooting. *NORMA* 13(2):86–102. <https://doi.org/10.1080/18902138.2017.1390658>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

