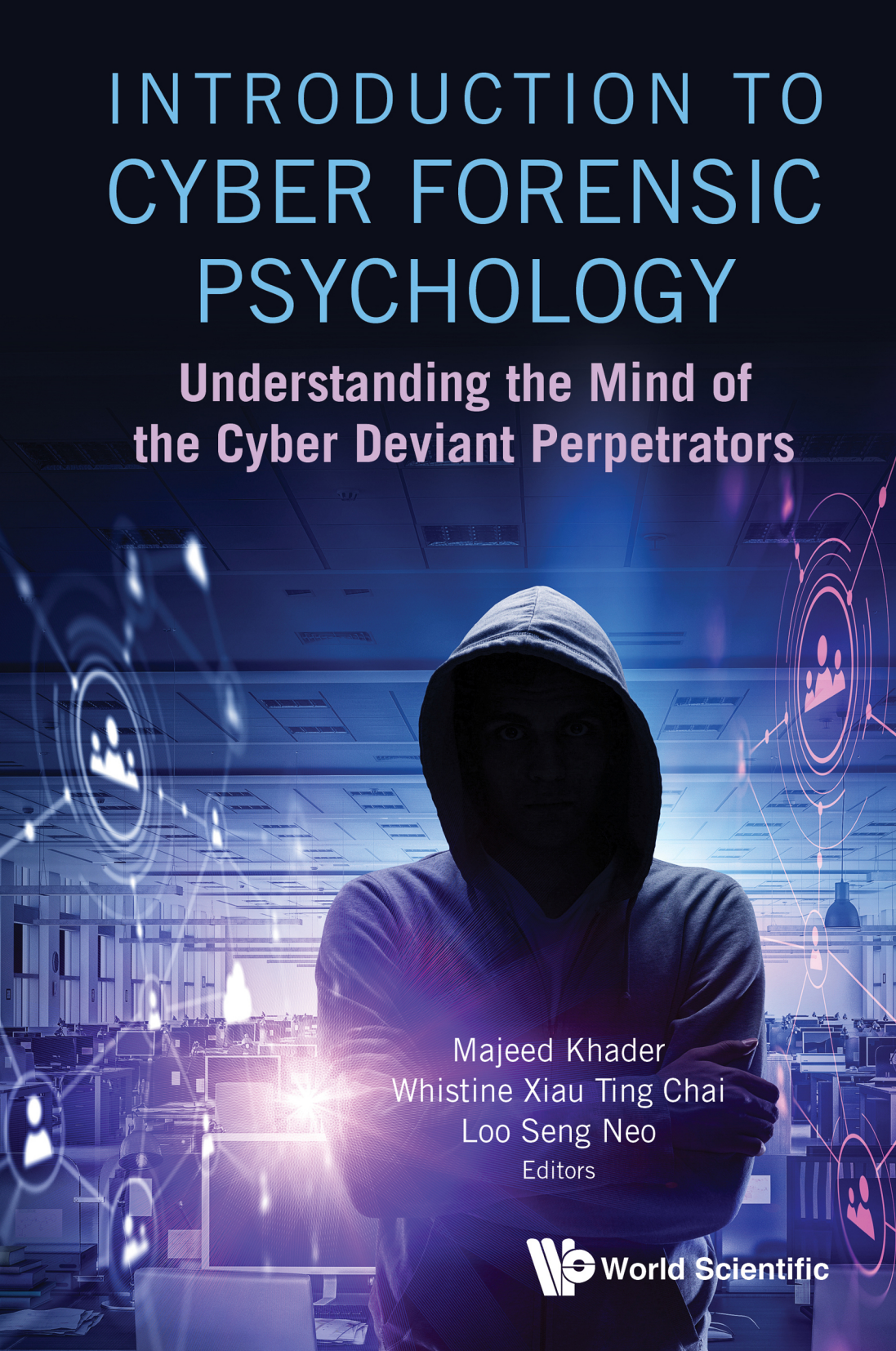
# INTRODUCTION TO CYBER FORENSIC PSYCHOLOGY

## Understanding the Mind of the Cyber Deviant Perpetrators



Majeed Khader
Whistine Xiau Ting Chai
Loo Seng Neo
Editors

World Scientific

# INTRODUCTION TO CYBER FORENSIC PSYCHOLOGY

## Understanding the Mind of the Cyber Deviant Perpetrators

This page intentionally left blank

# INTRODUCTION TO CYBER FORENSIC PSYCHOLOGY

## Understanding the Mind of the Cyber Deviant Perpetrators


Editors

Majeed Khader
Whistine Xiau Ting Chai
Loo Seng Neo

Home Team Behavioural Sciences Centre, Singapore



World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI • TOKYO

*Published by*

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
*USA office:* 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Control Number: 202193080

**British Library Cataloguing-in-Publication Data**
A catalogue record for this book is available from the British Library.

**INTRODUCTION TO CYBER FORENSIC PSYCHOLOGY**
**Understanding The Mind Of The Cyber Deviant Perpetrators**



ISBN 978-981-123-240-4 (hardcover)
ISBN 978-981-123-241-1 (ebook for institutions)
ISBN 978-981-123-242-8 (ebook for individuals)

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/12164#t=suppl

Desk Editors: Jayanthi Muthuswamy/Karimah Samsudin

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

https://doi.org/10.1142/9789811232411_fmatter

# Acknowledgements

It has been a privilege for us to work on this emerging but important topic on the psychology of cybercrime. The completion of this book would not be possible without the support of everyone who have contributed in one way or another. This book is a testament to the strong support we have received from our partners and colleagues working in the field of forensic psychology and cybersecurity. Therefore, we would like to acknowledge everyone who has played a critical role in the completion of this book.

It has been a remarkable journey for us and we would like to first express our most heartfelt gratitude towards our authors for their contributions to this book. A huge thanks to all our authors for your expertise, open-handedness, passion, and patience. It has been a great pleasure working with you and we are grateful for your faith in this project. We would also like to thank World Scientific Publishing, the publisher of this book, and in particular, Ms. Karimah and Ms. Jayanthi who have supported us tirelessly throughout this book project.

Next, we would wish to acknowledge those who have contributed to the various stages of the production process. A number of the Home Team Behavioural Sciences Centre (HTBSC) colleagues have graciously offered their time, insights, and administrative support to this project. We would like to acknowledge Mr. Ken Chen, Mr. Hou Minzheng, and Ms. Pamela Goh for their valuable assistance and reviews that have greatly improved the coherence and readability of the book. Ms. Crystal Koe and Ms. Phoebe Ng have also been a huge support in providing valuable inputs on the formatting and proof-reading issues during the final edit of the book. We would also like to thank Mr. Joel Ong for sharing his

research and Mr. Chan Chunmu for his kind administrative support. Last but not least, our sincere appreciation to all our colleagues at HTBSC for their personal, professional, and moral support towards this project.

In addition, we wish to highlight that this project is not an official government endeavour, and thus the views expressed here represent the views of the authors and editors only. The chapters of this book do not represent the official views of the Government of Singapore in any way. Notwithstanding that, we are grateful for our management's support and encouragement towards our work and research. We are also thankful for the support from Ms Chua Lee Hoong, our valued-partners from MHA (Joint Training Centre [JTC], Heritage Development Unit [HDU], Centre for Protective Security Studies [CPSS]), and Centre of Excellence for National Security (CENS). At Nanyang Technological University (NTU), our thanks go to Associate Professor Joyce Pang, Associate Professor Ringo Ho, and Associate Professor Kumar Ramakrishna for their support and guidance.

Finally, we would like to take this opportunity to thank the members of our families. Majeed is deeply thankful to Leong Tscheng Yee, Tasneem and Raouf, and his mum Hawa, for their unconditional love and support. Whistine is grateful to her family, her supportive husband Chua Sin Long, and loving children, Isabelle and Matthias, for their unconditional love. Loo Seng is deeply indebted to his wife, Onpapha, and his three children, Xi Zhen, An Qi, and An Ping, for their love and support. We truly appreciate their invaluable support and love.

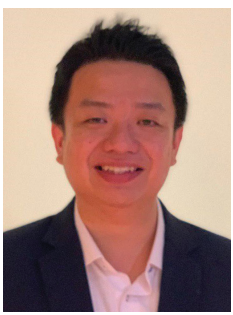
To our readers, thank you for taking the time to read this book. We hope you find our humble project a pleasure to read!

https://doi.org/10.1142/9789811232411_fmatter

# About the Editors

**Majeed Khader** is Chief Psychologist of the Ministry of Home Affairs (MHA), Singapore, and Director of the Home Team Behavioural Sciences Centre. He teaches criminal and forensic psychology as an Associate Professor (Adjunct) at Nanyang Technological University and Adjunct Honorary Associate Professor at National University of Singapore. For more than two decades, Dr. Majeed has overseen the development of psychological services in the areas of crime, terrorism, resilience, employee selection, deception psychology, leadership, crime profiling, and crisis psychology. For his work, he was awarded the National Day Public Administration Award (Bronze) in 2006 by the President of Singapore and the Public Administration Award (Silver) in 2014. He has been four times chair of the Asian Conference of Criminal and Operations Psychology, held in Singapore. He is also the Asian Director of the U.S.-based Society of Police and Criminal Psychology (SPCP). He is an active volunteer in several special needs institutions in Singapore including the Special Needs Trust Company.

**Whistine Chai Xiau Ting** is a Lead Psychologist with the Home Team Behavioural Sciences Centre. Whistine plays a pivotal role in conducting research and training for law enforcement officers in the Ministry of Home Affairs in the domain of forensic and investigative psychology. As Assistant Director of the Crime, Investigation and Forensic Psychology Branch, she has built numerous new capabilities to support operations and policymaking through the use of behavioural science approaches. Leading a team of psychologists and research analysts, Whistine also fronted multiple cross-agencies, as well as inter-ministerial and national forensic psychology research and programmes. A graduate of the National University of Singapore (NUS), she also holds a Master's degree (with Distinction) in Clinical Forensic Psychology from King's College London, where, she received the John Gunn Prize for Highest Overall Mark and the Sheilagh Hodgins Prize for Highest Dissertation Mark in the Forensic Mental Health programme. She also received the "Psychologist of the Year" award in MHA in 2020.



**Neo Loo Seng** is a Principal Behavioural Sciences Research Analyst with the Home Team Behavioural Sciences Centre at the Ministry of Home Affairs, Singapore. For the past 13 years, Loo Seng has been leading a team of research analysts and interns to research on emergent trends on terrorism, resilience, misinformation, and intergroup conflict, and distilling useful insights for the ministry through research reports as well as training for law enforcement officers. He has also trained grassroots leaders and the general public about the threat of terrorism and how to prepare for the 'day after' terror. Based on his research, Loo Seng has presented at many international conferences and published many peer-reviewed journals and book chapters. He has also co-edited several books. He has achieved several awards such as the Research of the Year (2018). His research interests are online radicalisation, misinformation, and online threat assessment. Loo Seng is currently pursuing his PhD in psychology at the Nanyang Technological University (NTU).

https://doi.org/10.1142/9789811232411_fmatter

# About the Contributors

**Afreen Chawla** is a Psychologist from the Crime, Investigation, and Forensic Psychology branch of the Home Team Behavioural Sciences Centre (HTBSC), Ministry of Home Affairs (MHA). Afreen graduated from the National University of Singapore (NUS) with a Bachelor of Social Sciences (Hons.) in Psychology, with a minor in Communications & New Media and a second minor in Forensic Science. Her current projects cover the domains of scams and crime prevention, as well as psychological profiling and mental health. Her interests include criminal and clinical psychology, comprising the spectrum of mental health disorders and their treatment alongside their forensic implications. Besides these, she is a fan of all things music, enjoys baking and reading, and learning about everything from astronomy to archaeology to anatomy to art.

**Benjamin Ang** is Senior Fellow and Coordinator of the Cyber Homeland Defence Programme in the Centre of Excellence for National Security (CENS), an independent policy research institute in the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. He leads the team researching national security aspects of cybersecurity, disinformation, influence operations, surveillance, data privacy in artificial intelligence (AI), and technology policy. His varied qualifications include Law, Business Administration, and Management Information Systems (MIS). He spent 20 years in the private sector, at various times as a lawyer, IT Director, and Legal Technology Consultant (including digital forensics), before entering academia.

**Benjamin Goh** is a Global Information Assurance Certification (GIAC) Certified Incident Handler, and holds a Master in Public Policy from the John F. Kennedy School of Government. He is a cybersecurity specialist working in Singapore's Government Technology Agency's Cybersecurity Group.

**Chan Meng Fai** is a Certified Information Systems Security Professional (CISSP), and has worked in cybersecurity roles across government, private sector, and education. He is a cybersecurity specialist working in Singapore's Government Technology Agency's Cybersecurity Group.

**Hou Minzheng** is a Doctoral candidate in Social Psychology at the National University of Singapore (NUS), and a recipient of the President's Graduate Fellowship. His core research area resides in motivation and decision-making. A member of the Situated Goal Pursuit Lab, he is also actively involved in research projects aimed at enhancing intergroup relations based on motivation science.

Prior to pursuing his PhD, Minzheng was at the Institute of Policy Studies (IPS), Lee Kuan Yew School of Public Policy (LKYSPP), where he spearheaded national level surveys on social resilience. Minzheng also served as a Naval Officer in the Republic of Singapore Navy for more than seven years where he was actively deployed for frontline operations and intimately involved in operational policy development.

Minzheng obtained his Bachelor's degree (highest honours) in Psychology and Economics from the University of Michigan, Ann Arbor. During his time at Michigan, he received the prestigious Goldstein Prize (Marshall Sahlins Social Science Award) as a recognition of his excellent academic and research achievements.

**Jeffery Chin** has been a staff Psychologist with the Home Team Behavioural Sciences Centre (HTBSC) since 2006. Incepted in 2006, the HTBSC is an applied behavioural sciences research, training, and operational support unit in the Ministry of Home Affairs (MHA), Singapore. One of the Centre's key remit is to translate knowledge from the field of psychology and the behavioural sciences into operational knowledge that law enforcement, safety, security and correctional agencies under the auspices of MHA may utilise or incorporate into their policies, operations and practice.

As one of the pioneers at the HTBSC, Jeffery played an important role in setting up the unit in its initial years and oversaw its development from a programme (the Behavioural Sciences Programme) in 2006 to a full-fledged psychological research and training centre (the Home Team Behavioural Sciences Centre) presently. He has been involved in roles and projects across several domains that supported the operations and development of Singapore law enforcement and security officers over the years. The roles and projects include crime and investigation support research, training and consultation, crisis intervention and resilience management, crisis negotiation operations and leadership development and assessment.

**Jessie Janny Thenarianto** is a Behavioural Sciences Research Analyst at the Home Team Behavioural Sciences Centre (HTBSC). She graduated summa cum laude from Universitas Ciputra, Indonesia, with a bachelor's degree in Psychology. Her current research interests include crisis, resilience, internet, and crime. Some of her published works include the social media response after the 2016 and 2017 Jakarta bombings in *Learning from Violent Extremist Attacks: Behavioural Sciences Insights for Practitioners and Policymakers* [2018] and political symbolism during elections in *Psikologi dan Integrasi Bangsa: Seri Sumbangan Pemikiran Psikologi untuk Bangsa 4* (2020).

**Joey Low** is a Correctional Rehabilitation Specialist (CRS) with the Singapore Prison Service (SPS). As a Specialist employed in the after-care setting, she provides case management services and support offenders' rehabilitation and reintegration into the society upon their release. Joey has passion in the field of Forensic and Correctional Psychology. During her undergraduate studies in Nanyang Technological University (NTU), she went through a three-month stint as an intern with the Rehabilitation Evaluation Branch of SPS, where she had the opportunity to conduct an evaluation on the effectiveness of the rehabilitation programme. Prior to which, she undertook an internship with the Home Team Behavioural Sciences Centre (HTBSC), where she underwent an in-depth research on the topic of "Command Leadership: Managing Information during Crisis," and presented her findings to the Home Team officers. In collaboration with HTBSC, Joey and her team constructed a checklist on the "Traits of a Good Liar" as part of her Final Year Project. The team was offered an opportunity to present their findings at the

Asian Conference of Criminal & Operations Psychology (ACCOP). Joey graduated from NTU with a Bachelor Degree in Psychology (First Class Honours), she also holds a Diploma in Psychology from Temasek Polytechnic.

**John Yu** is a Psychologist from the Crime, Investigation, and Forensic Psychology branch of the Home Team Behavioural Sciences Centre (HTBSC). John graduated from the National University of Singapore (NUS) with a Bachelor of Social Sciences (Hons.) in Psychology. He currently carries out research in the areas of criminal profiling and emerging criminal and deviant behaviours of interest and concern. He has conducted training for law enforcement personnel on topics in criminal and forensic psychology. John is also a co-organiser of the Criminal Behavioural Analysis Competition in Singapore, an annual nationwide competition that introduces tertiary students to criminal profiling and investigative techniques, psychological first aid, and crisis negotiation skills.

**Karthigan Subramaniam** is a Behavioural Sciences Research Analyst with the Home Team Behavioural Sciences Centre (HTBSC) at the Ministry of Home Affairs (MHA), Singapore. Karthigan's research interests and expertise fall within the area of investigative interviewing, with particular attention to the following themes: rapport, interviewing techniques, information elicitation and memory. In addition to research, he also conducts training for law enforcement officers. Karthigan obtained a summa cum laude (i.e., the highest honours) undergraduate degree in Psychology from the University at Buffalo, State University of New York, and is also a member of the Phi Beta Kappa Honours Society.

**Muhammad Faizal B Abdul Rahman** is a Research Fellow with the Centre of Excellence for National Security (CENS) at the S. Rajaratnam School of International Studies (RSIS). He holds a Bachelor of Business Administration (with Merit) from the National University of Singapore (NUS). He completed his Master of Science in Strategic Studies at RSIS, specialising in terrorism studies. His dissertation examined the grand strategies of Al Qaeda and the Islamic State (Daesh), focussing on asymmetric warfare and cities as a jihadi battlespace. Faizal previously served with the Singapore Ministry of Home Affairs, where he was a Deputy Director and had facilitated international engagements with

foreign security counterparts. He also had postings in the Singapore Police Force where he supervised and performed intelligence analysis, achieving several commendation awards including the Minister for Home Affairs National Day Award (2009) for operational and analysis efficiency; and in the National Security Research Centre (NSRC) at the National Security Coordination Secretariat (NSCS), where he led a team to research emergent trends in domestic security and monitor terrorism-related developments. His research interests are counter-terrorism, strategic foresight and intelligence, and implications of global trends on homeland security.

**Nur Aisyah Abdul Rahman** is a Behavioural Sciences Research Analyst at the Home Team Behavioural Sciences Centre (HTBSC). She has an undergraduate degree in Psychology. Aisyah's areas of research involve social cohesion issues in light of violent extremism (e.g., prejudice, Islamophobia), microaggressions, and right-wing extremism. She is part of a team that regularly conducts seminars and trainings for police officers, and religious and community leaders. She enjoys learning more about violent extremism and related social phenomenon. She has also published and contributed to several reports, journal articles, and book chapters on topics related to violent extremism and social cohesion, as well as presenting her research findings at various conferences and seminars.

**Omer Ali Saifudeen** is currently serving as Senior Assistant Director at the National Security Research Centre (NSRC), National Security Coordination Secretariat (NSCS), Prime Minister's Office (PMO).

**Pamela Goh** is a Behavioural sciences research analyst with the Home Team Behavioural Sciences Centre (HTBSC), Ministry of Home Affairs (MHA), Singapore. Her research experience and repertoire in the organisation includes building national resilience, both offline and online, in relation to nationwide crises. Some of her past work included crowd behaviours during crises, building social cohesion, crisis communications, and cyber hygiene. Additionally, she has also delved into the area of whole-of-society engagement in response to crises, having had organised two roundtables involving academics, governmental officials, non-governmental organisations, and other law-enforcement agencies, to discuss about community responses to terror attacks. Pamela is also currently pursuing her PhD at the Nanyang Technological University

(NTU), Singapore, with an avid interest in community resilience and cohesion in times of crises. In part of her dissertation, she is looking at understanding whether and why people would act for themselves or for others during times of threatened survival, such as a terror attack.

**Shannon Ng** is a Behavioural Sciences Research Analyst with the Home Team Behavioural Sciences Centre (HTBSC) at the Ministry of Home Affairs (MHA), Singapore. Shannon graduated from the Nanyang Technological University (NTU) with a Bachelor of Arts in English with Honours (Distinction) and a minor in Psychology. Her current field of research covers cyber related crimes and violent crimes—in their emergence, deviancy and prevention/interception. Her interests lie in criminal profiling and youth offending, and how power dynamics of race, age, gender, social class, etc., contributes to crime and its solutioning. She also uses her flair for the arts and experience in social media public engagement in the creation of a set of vibrant posters for the annual Criminal Behavioural Analysis Competition in Singapore. Other than that, she is probably busy rearranging her room, collecting rainwater for her plants, doing some craft or catching up on Netflix.

**Stephanie Chan** is a Senior Psychologist at the Home Team Behavioural Sciences Centre (HTBSC). She holds a master's degree in forensic psychology from the University of Leicester (UK), and has relevant experience in crime and forensic psychology. Her research interests include cyberstalking, cyber-deviant behaviours, and the detection of deception for law enforcement, intelligence gathering, and border control purposes.

**Vivian Seah** is a Psychologist with the Crime, Investigation, and Forensic Psychology (CIFP) Branch of the Home Team Behavioural Sciences Centre (HTBSC), Singapore. Key areas of her work at HTBSC include crime and behavioural analysis and behavioural sciences research on scams, drugs and organised crime. Her research interests also include the detection of deception and investigative interviewing. She organises and conducts trainings for Home Team officers in the area of crime and behavioural analysis. Through her work in HTBSC, she engages the community to build interest in the area of investigative and forensic psychology by being part of the organising committee of the annual Criminal Behavioural Analysis Competition (CBAC) and the Asian

Conference of Criminal and Operations Psychology (ACCOP) 2019. She is a trained volunteer Victim Care Officer (VCO) with the Singapore Police Force, where she provides psychological support to victims of crimes. She is also a psychologist with the Singapore Police Force Crisis Negotiation Unit, where she is involved in crisis negotiation operations by giving psychological inputs during negotiations. Academically, Vivian holds a Bachelor of Social Science in Psychology (First Class Honours) and a minor in Education Studies from the Nanyang Technological University (NTU), Singapore.

**Xingyu Ken Chen** is a Senior Behavioural Sciences Research Analyst at the Home Team Behavioural Sciences Centre (HTBSC). Some of his other writings include psychological vulnerabilities to fake news in the aftermath of a terror attack in *Learning from Violent Extremist Attacks: Behavioural Sciences Insights for Practitioners and Policymakers* (2018), the intersection of crime and fake news in *Encyclopedia of Criminal Activities and the Deep Web*. His current research interests include online misinformation, information operations, strategic communications and natural language processing.

**Yasmine Wong** is a Senior Analyst with the Centre of Excellence for National Security (CENS) of the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore. Yasmine's current research focuses on issues pertaining to social resilience, social cohesion and inter-group relations.

This page intentionally left blank

https://doi.org/10.1142/9789811232411_fmatter

# Contents

This page intentionally left blank

# Part 1

# Introduction

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0000

# Introduction to Cyber Forensic Psychology

Whistine Chai Xiau Ting*, Shannon Ng†, and Neo Loo Seng‡

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Chai_xiau_ting@mha.gov.sg

†shannon_NG@mha.gov.sg

‡neolooseng@gmail.com

> This is not a race that we can win alone. A rising tide raises all boats, and collaboration between governments, businesses, academia, and individuals across multiple fronts will be a crucial component to our success. Moving forward, it is important not just to build defences for the cyber-threats of today, but also to promulgate the infrastructure, capabilities, mindsets and friendships that will enable us, as an international community, to tackle the cyber threats of tomorrow.
>
> —David Koh, CE of the Cyber Security Agency of Singapore (CSA) [Koh, 2018]

## 1 The Blooming Internet and Technological Ecosystem

The world is increasingly interconnected, evermore fast-paced, and we are caught in a web of our making. The World Wide Web was first created in

1989 with the launch of the first website in 1991. Within 28 years, the number of sites has exploded to a staggering 1.72 billion [Armstrong, 2019]. Not only are we talking about an ever-expanding cyberspace that we are still playing catch-up in regulating, we are also faced with a larger online population that Cybersecurity Ventures predicts will grow to 75% of the projected world population in 2022 and 90% of that in 2030 [Morgan, 2019].

As of January 2019, the worldwide internet penetration rate was at 57% (4.38 billion internet users), having grown by 9% (367 million internet users) since January 2018 [Hootsuite, 2019]. While the internet penetration rate for Southeast Asia was at 63%, Singapore numbered at 84% with 4.92 million internet users within her shores. Notably, the high functionality of the internet means that our daily lives will only become more integrated with the cyberspace.

Taking the COVID-19 pandemic as an example, digitalisation and technological advancements have further accelerated and deepened our reliance on the internet. We see organisations shifting towards having their employees work remotely via remote access, and institutions having their students learn from home via home-based learning programmes. Even religious meetings were carried out on online platforms, forcing the elderly and the non-technological savvy in these communities to engage with the internet on a deeper level than what they were used to. The usefulness of the internet in maintaining as much semblance of normalcy during such a crisis cannot be refuted. We can only expect businesses and other sectors of life to embrace technology and cyberspace even more so going forward post-COVID-19.

As different aspects of our lives become increasingly intertwined with cyberspace, we also see another trend—a surge in the amount of online personal data on individuals. This data is quantified as four zettabytes in 2016, and is set to reach to 96 zettabytes by 2020 [Morgan, 2019], multiplying twenty-four times in just four years. Ninety-four percent of this mass of data is predicted by Cisco [2018] to be housed by cloud data centres by 2021, with only 6% still processed by traditional data centres. While the population of netizens grows closer to fully representing the entire human population, we cannot deny that the cyberspace is as real a space as the physical world we live in. Since safe practices in protecting our homes is second nature, the same applies to the protection of our personal data and loved ones online.

## 2 The Dark Side of the Cyberspace

Along with all the good that the internet brings, comes individuals who seek to exploit its vulnerabilities. Cybercriminals are defined as individuals who use technology to commit malicious acts of cybercrime on digital systems or networks to profit from stolen company or personal data [Trend Micro, n.d.]. Cybercrime is expected to cost the world $6 trillion annually by 2021 [Morgan, 2019]. This cost covers diverse forms of cybercrime including theft of money, intellectual property and data, the destruction of data, embezzlement, fraud, and business disruption. Similarly, International Business Machines Corporation (IBM) Security [2019] reported the increased average total cost of a data breach in 2019 to be US$3.92 million.

This surge in cost does not come as a surprise as we see a surge in netizen population due to the increase in global internet penetration rate. With a burgeoning netizen population comes an inflation of the number of smart devices that communicate wirelessly and further balloons the amount of online data. Naturally, we expect this shift to also correspond with an increased number of digital targets for cybercriminals. Cybercrime would eventually cut through every strata of urban society, and become pervasive across all age groups or social classes owing largely to the Internet of Things (IoT).

The Internet of Things (IoT) refers to the network of devices that can communicate wirelessly or access the internet [Ranger, 2020]. It is made up of IoT devices which include any item ranging from a smart lightbulb to a driverless vehicle, as long as it can be connected to the internet. Unfortunately, the strength of IoT is also its greatest weakness. IoT's ability to wirelessly communicate with each other makes it susceptible to attacks by remote access. Perpetrators no longer need physical access to our IoT devices in order to breach security to get to our data or invade our privacy. For instance, while Singapore held classes online via the video-conferencing platform, Zoom, during the COVID-19 quarantine period, a class was 'Zoombombed', in which hijackers flashed pictures of penises to a class of 39 secondary one girls, instructing them to "show [them their] boobs" [Baharudin, 2020]. This was not an isolated incident as there were international reports of Zooombombing even on password-protected chats. Clearly, IoT devices are entryways to our valuable data. More IoT means opening more access points to our data and as a result, more security is required to safeguard them.

Online data is valuable to cybercriminals as they could be 'weaponised' to cause harm to persons, organisations, and countries. Criminal use of such data is problematic because it removes the autonomy of victims and allow the perpetrators to exploit the information gained to achieve wealth, power, or other desires they may have. The WannaCry ransomware is one of the most, if not the most, prolific example of cybercrime the world has experienced. In 2017, it affected 150 countries, over 250,000 systems, and demanded a ransom of $300 in bitcoins per system [Fruhlinger, 2018]. While not all victims had paid the ransom, they still suffered loss; the National Health Service in the United Kingdom, for example, lost £92 million in cancelled appointments and subsequent necessary upgrades to their security systems [Field, 2018]. Beyond external cyberattacks, organisations need to be vigilant against internal threats and vulnerabilities such as WikiLeaks data breach by Bradley Manning. Most of all breaches involve someone unknowingly compromising systems and company data from the inside as a result of poor cyber hygiene, malicious insiders, and negligence [Vishwanath *et al*., 2020].

It is important at this point to recognise that the dark side of cyberspace does not solely revolve around gaining illegal access to computer networks to gain monetary incentives and online data. There are lots of other examples of how individuals with malicious intent have reorganised their operations online. They range from terrorists who use the internet to influence and radicalise new members, to operators of information operations that manipulate online discourse on social media during elections, to criminals who livestream their acts of violence such as the Uppsala gang-rape in Sweden, to online expressions of hate by netizens that lead to violent protests as seen in the 2018 Sri Lanka Kandy Riots, and to individuals who engage in deviant online behaviours such as the circulation of upskirting photos. While these individuals may not be motivated by money (unlike the conventional profile of cybercriminals that we are familiar with), what is certain is that the internet has played an imperative role in the way malicious activities are being conducted.

## 3 The Need to Regulate and Control the Cyberspace

As the world witnesses an upward trend of such crime and security concerns in the online sphere, the need to regulate and control cyberspace

will continue to be of utmost importance to the entire world. French philosopher Michel Foucault theorised that "knowledge and power are integrated with one another" and that it is "impossible for knowledge not to engender power" [1980, p. 52].

Indeed, the internet and wireless access that permeate our world threatens to transfer knowledge and power to a much wider range of people than previously seen. The Cambridge Analytica and Facebook scandal is a clear illustration of how that knowledge and power can be transferred from an individual to another. As many as 87 million Facebook users had their information sold without their knowledge to Cambridge Analytica for the purpose of influencing politics [Lapowsky, 2019]. This incident speaks volumes about the need to regulate cyberspace and the companies that operate on them. Furthermore, the real-life consequences (acts of violence, communal disharmony, disruptions to the lives of victims) associated with cybercrimes place additional responsibility on law enforcement agencies to respond with the appropriate technological interventions. These security agencies are compelled to transform the way they identify potential persons-of-interest, collect usable intelligence, and conduct threat assessments.

It is with this understanding that we look on to the existing measures employed locally to curb the unruliness of the cyberspace.

## 3.1 *Local measures to counter cybercrime*

In Singapore, cybersecurity is approached as an internal and international phenomenon. As such, her plans dive into immediate necessary steps to safeguard the present while simultaneously engaging in measures that will yield future benefit. This comes in the form of the nation beefing itself up internally as well as growing international relations, and engaging in the global discourse of regulating cyberspace as a member of the United Nations Group of Governmental Experts (UN GGE).

In the past, Singapore has dealt with a series of cyberattacks, namely the 2013 attacks by 'The Messiah' ["Top 5 biggest cyber," 2017], followed by a more serious attack by a hacker group named Whitefly on SingHealth, in which personal data, including that of Singapore's Prime Minister, was stolen [Kwang, 2019]. From 2018 to 2019, the total number of reported criminal cases in Singapore increased by 6.3%, most of which could be attributed to a significant increase in online scam cases. We can understand just how significant cybercrime is when we consider how

reported crime rates in Singapore would fall by 4.6% if we had removed all reported online scam cases [Singapore Police Force, 2020].

Reported scam cases have increased by 53.5% since 2018 with the top scams being e-commerce scams, loan scams, and credit-for-sex scams. The Singapore Government has been actively looking for solutions to mitigate and combat scams, which birthed the Anti-Scam Centre (ASC) on 18 June 2019. A new inter-ministry taskforce comprising the Ministry of Home Affairs (MHA), Ministry of Communications and Information (MCI), and Ministry of Trade and Industry (MTI) was also formed to explore the solutions in reducing scams and other fraudulent online activities [Wong, 2020].

Another of Singapore's approach in countering cybercrime is the establishment of the Cyber Security Agency of Singapore (CSA) on 1 April 2015. This is probably the main pillar in the nation's fight against cybercrime. CEO of CSA, David Koh, outlined the four facets of Singapore's approach in his speech at the Third Annual Billington International Cybersecurity Summit [Koh, 2018, paras. 12–15]:

(1) Build resilient infrastructure to ensure continual provision of essential services;
(2) Go beyond Critical Information Infrastructure sectors to engage businesses and individuals to develop a safe and secure cyberspace;
(3) Build a vibrant cyber security ecosystem; and
(4) Strengthen international partnerships.

Singapore has adopted a pragmatic approach to inoculate her citizens against cyber threats. Radical moves such as segregating the internet and the government's internal networks have been made with favourable yield [Koh, 2018]. Staying true to her nature, plans for emergencies have been developed and practiced, while efforts to educate individuals and small businesses on cyber safety and security to improve their cyber hygiene have already begun with campaigns like the "Go Safe Online 2019" [CSA, 2019]. In balancing the ominous prospects of cybersecurity, Singapore also acknowledges the economic benefits it brings for the nation. In recognition of this, Singapore has started a $190 million research and development (R&D) programme for national cybersecurity [Koh, 2018], pumped in resources to grow local cybersecurity companies like Swarmnetics and Phishnow ["Singapore's Cybersecurity startup," 2020], as well as attracted top international cybersecurity firms

like Symantec and Trend Micro to set up branches in Singapore [Amran, 2019].

## 4 Moving Beyond Technological and Legal Solutions: The Role of Behavioural Sciences

While we acknowledge the need for the right technological and legal approaches, equal emphasis should be placed on the use of behavioural science and psychology as a combative tool in the matter, and with good reason.

Extant research shows differences in certain statistics and suggestions in countering cybercrime, yet a common reiteration is consistent among experts—the fact that humans are the weakest link [Bissell *et al*., 2019]. Research reported more than 90% of successful attacks and data breaches come from phishing emails that trick recipients into downloading malware or forwarding information to unauthorised actors [Morgan, 2019]. Studies also claim that open-source digital footprints (i.e., online behaviours on social media and internet) can be harnessed to better identify and understand potential cyber threats [Neo, 2020]. It is within these digital footprints that a potential cybercriminal's intention and warning signs may manifest, which can in turn be used to assess the threat they pose. Thus, the common underlying theme is the need to better understand the human actors and their cognitions and behaviours.

Recalling Foucault's assertion about knowledge and power, the knowledge we glean about the motivations and thought processes of cybercriminals would give us the edge we need to create more appropriate technological solutions to tackle cybercrime. In view of this, the insights from behavioural science, specifically cyber forensic psychology, would enhance our understanding of the factors that give rise to cyber deviancy and help answer key questions such as what does it mean to be criminal in the cyberspace? What are the underlying motivations behind cybercrime and deviant acts, and most importantly, what can we do to mitigate such acts? For the purpose of this book, cyber forensic psychology is defined as the study of the mind and behaviour of cybercriminals, which examines common forensic psychology research domains such as offender profiling, risk and protective factors of criminality, and strategies to minimise offending.

This book will comprise a range of chapters exploring two general themes: (1) emerging cybercrimes and cyber enabled crimes; as well as (2) strategies for assessment, prevention, and interventions. For the former, the chapters concentrate on elucidating the psyche of cybercriminals who engaged in (1) violence and deviant online behaviours; (2) sexual and deviant online behaviours; (3) online hate crimes and information operations; (4) cyber scams. For the latter, the chapters focus on how law enforcement can learn and identify strategies to; (5) intervene and mitigate cyber threats; and (6) identify future concerns regarding cybersecurity threats.

There are two ways by which this book adds unique value to the extant literature on cybersecurity. Firstly, this book encapsulates an endeavour to solicit and harness the insights of practitioners, policymakers, and subject-matter experts in the field of cybersecurity. Slated for practitioners and policymakers, this book allows contributors to share their perspectives and wealth of experience, all of which could enrich and serve as a critical and timely resource to deepen our knowledge about the emerging issues as well as strategies to combat them. Secondly, this book aims to provide insights from a Singaporean perspective such that the information presented will be relevant to local practitioners and policymakers. To this end, the editors (from the Home Team Behavioural Sciences Centre, Ministry of Home Affairs) have invited local partners to contribute chapters on the topic of cybersecurity, where much of the scholarship has originated from a Western perspective. More importantly, the editors seek to convey a spirit of openness, wherein the contributors are strongly encouraged to share any useful insight that has not yet been heard of or which runs contrary to prevailing conventional views. Instead of echoing conventional views, this approach acknowledges the rich and varied range of disparate questions and issues associated with cybersecurity, and serves as a starting point for readers to reflect and envision new areas of research.

## 5 Part 2: Cyber Crimes and Cyber Enabled Crimes

To better understand the mind of cybercriminals, emerging issues of cybercrime and cyber-enabled crime will be examined. Cyber-enabled crimes are ‘traditional’ crimes that are adapted for and committed on cyberspace (e.g., sexual abuse, scams), whereas cybercrimes are crimes

that target computer networks, hardware, or software (e.g., Ransomware). The chapters in this part have been organised broadly into the following sections:

- Section A: Violent and Deviant Behaviours Online (Chapters 1 to 3)
- Section B: Sexual and Deviant Behaviours Online (Chapter 4)
- Section C: Hate Crimes Online (Chapters 5 to 8)
- Section D: Cyber Fraud and Scams (Chapters 9 to 10).

## 5.1 *Violent and deviant behaviours online*

The first three chapters examine the online behaviours of cybercriminals that are associated with violence in the real world. Whistine Chai Xiau Ting and John Yu's opening chapter on 'Influence of Social Media on Deviant Acts: A Closer Examination of Live-streamed Crimes' (Chapter 1) examine an emerging phenomenon in which individuals would broadcast their criminal activities over social media. They discuss the motivations underlying live-streamed crimes as well as how law enforcement should mitigate and combat live-streamed crimes. Stephanie Chan, in her chapter 'Hidden in the Shadows: Stalkerware Usage in Intimate Partner Stalking', (Chapter 2) focuses on the use of smartphone spyware in the context of intimate partner stalking, and examines the prevalence of stalkerware, the perpetrator's mindset in using stalkerware, and recommendations to alleviate the harm caused to victims. On a related note, John Yu, in his chapter 'Digital Self-Harm: A Peek into the Mind of an Online Self-Aggressor' (Chapter 3), provides a probable glimpse into the mind of one who has committed digital self-harm by reviewing the limited literature and drawing possible associations within other related domains, such as the psychology of cyberspace and the cyber-self, bullying, and self-harm.

## 5.2 *Sexual and deviant behaviours online*

In recent times, digital technologies are increasingly used as tools of harassment, abuse, violence, and sexual deviance. Bearing this in mind, Vivian Seah, in her chapter 'Cyber Sexual Deviance: Delving into Image-Based Sexual Abuse' (Chapter 4), builds upon past research on Image-Based Sexual Abuse (IBSA) from sociological and feminist-criminological angles, and delves into the prevalence, underlying motives, and impact

of IBSA. She then introduces several mitigation strategies for the issue of IBSA.

## 5.3 *Hate Crimes Online*

The central question underpinning Section C of the book revolves around how individuals with malicious intentions have exploited the affordances of the internet to cause social disharmony. Nur Aisyah, in her chapter 'Is Technology Making You Prejudiced? How Technology is Enabling Hate IRL' (Chapter 5), examines four characteristics of the online environment that cyber perpetrators have exploited to spread hate. In it, she discusses various systemic and human interventions that law enforcement can adopt to combat cyberhate. In the chapter 'Rebellion Against the State: A Social Perspective on How the Online Space Fuels Collective Action' (Chapter 6), Hou Minzheng seeks to answer the question of how the online sphere can exert a profound impact on driving collective action against governments. Leveraging on various psychological theories, he uncovers the role of the internet in influencing individuals' identity and enhancing their motivational drive in collective action participation. Next, Yasmine Wong, in her chapter 'Victim and the Cyber Vigilante: An Additional Perspective on Cyber Vigilantism' (Chapter 7), examines the intricacies associated with cyber vigilantism. In it, she examines why victims of sexual harassment and assault might use doxing against their perpetrators, and explores the cascading mobilisation of the public in this context, which could lead to collective participation in cyber vigilante behaviours. Finally, Xingyu Ken Chen and Jessie Janny Thenarianto, in their chapter 'Understanding the Growing Prevalence of Information Operations on Social Media' (Chapter 8), look at how information operations actors have succeeded in manipulating online public opinion via social media during crucial times, such as elections. In particular, they share two approaches that authorities can adopt to limit the effectiveness of information operations actors.

## 5.4 *Cyber fraud and scams*

In the final section of part two, attention is directed towards cyber enabled crimes that plague Singapore. Jeffery Chin, in his chapter 'Love Cheats: The Psychology of Love Scams' (Chapter 9), provides a psychological perspective of the key issues underlying love scams. He identifies the

factors that make them work, as well as the impact of love scams on its victims. Leading on from this, Afreen Chawla, John Yu, and Shannon Ng, in their chapter 'Cybercrime and scam amidst COVID-19' (Chapter 10), consolidate the known tactics employed by cybercriminals during the COVID-19 crisis. They share how certain human attributes may increase one's risk of victimisation, and suggest strategies to circumvent it.

## 6 Part 3: Assessment, Intervention, and Prevention

Part 3 of this book will focus on key solutions that can be used to combat emerging cybercrimes and cyber enabled crimes. The chapters in this part have been organised broadly into the following sections:

- Section E: Insights for Assessment, Prevention, and Intervention (Chapters 11 to 14)
- Section F: The Future of Cybersecurity (Chapters15 to16)
- Section G: Special Chapter (Chapter 17).

### 6.1 *Insights for assessment, prevention, and intervention*

Section E of the book contains four chapters on measures that can mitigate the threat of cybercrime. Benjamin Ang, in his chapter 'Legal Issues and Ethical Considerations in Cyber Forensic Psychology' (Chapter 11), explores the legal and ethical considerations in applying forensic psychology in cybercrime prevention, detection, and response. Specifically, he raises practical considerations for researchers who are conducting cybercrime related research, as well as for investigators who may use cyber tools in their line of work. Complementing the previous chapter, Chan Meng Fai and Benjamin Goh, in their chapter 'Optimise Defender's Advantage: Practical Approaches for Cybersecurity Defence' (Chapter 12), outlines the challenges facing cybersecurity practitioners, and proposes solutions they can adopt to mitigate the seemingly impossible risk associated with cybersecurity. Next, Omer Ali Saifudeen, in his chapter 'Hacking the Hacker's Psyche' (Chapter 13), uses multiple case studies to illustrate the profile of a hacker, how they see themselves, their attitudinal characteristics, as well as their motivations. The insights gleaned can aid in the assessment of potential hackers. Finally, Pamela Goh, in her chapter 'Humans as the Weakest Link in Maintaining Cybersecurity: Building Cyber Resilience in Humans' (Chapter 14),

articulates the reasons why humans are the weakest link in cyber security, and outlines several measures that law enforcement can adopt to design human-centric approaches to combat cyber threats.

## 6.2 *The future of cybersecurity*

The two chapters under Section F adopt a futurology perspective and examine potential cybersecurity repercussions that we may face in the future. Muhammad Faizal, in his chapter 'Smart Homes: Where Rogue AI and Robots could Impair Security' (Chapter 15), discusses key points on cyber-attacks involving AI and robots in smart homes. He then proposes two defence approaches that authorities should consider in their efforts to ensure that smart homes are not compromised. On a related note, Karthigan Subramaniam, in his chapter 'Understanding and Mitigating the Risk of Hackercide' (Chapter 16), calls attention to the phenomenon of committing murder via IoT devices as a form of cybercrime and attempts to provide a realistic assessment as to why such an incident has not yet occurred. He also highlights three measures that can help to further mitigate this risk.

## 6.3 *Special chapter on sexting in Singapore*

The last section of the book is dedicated to an empirical research on sexting behaviours in Singapore. Joey Low and Majeed Khader (Chapter 17) introduce the concept of sexting, and highlight its prevalence in Singapore. They derive a list of psychosocial risk and protective factors based on a literature review, and then utilise Structural Equation Modelling to assess the relationship between the identified factors, intention to sext, and sexting behaviours. They opine that more attention should be given to looking into this phenomenon as sexting can lead to negative real-world consequences.

## 6.4 *Concluding chapter*

The last chapter titled 'The Future of Cyber-Forensic Psychology: How to Prepare' by Majeed Khader (Chapter 18) brings the discourse on cyber forensic psychology and its relevance for understanding cybercriminals and cybercrimes to a close by summarising the key learning lessons from the preceding chapters, as well as highlighting future research directions.

## 7 Conclusion

In conclusion, this book centres on the examination of cybercrime from a cyber forensic psychology perspective. It seeks to elucidate the psyche of cybercriminals and emphasise the importance of this perspective in mitigating any forms of cyber threats. Having said that, it is essential to also situate this understanding among other perspectives such as the technological and legal aspect, so as to achieve a more holistic understanding of the pressing problem of cyber security. Bearing this in mind, the various chapters can be seen as attempts to advance the topic and contribute to the endeavour of developing better cybersecurity policies and operational interventions. It is hoped that this book may serve as a starting point for greater discussions and study of this topic, as well as pave the way for many more outputs that can provide further insights for practitioners and policymakers.

## 8 Acknowledgement

The views expressed in this chapter are the authors' only and do not represent the official position or view of the Ministry of Home Affairs.

## 9 References

Amran, A. (2019, October 14). Top 43 Outstanding Cyber Security Companies in Singapore. *It.com.sg.* https://it.com.sg/services/cyber-security-companies/

Armstrong, M. (2019, October 28). *How many websites are there?* Statista.com. https://www.statista.com/chart/19058/how-many-websites-are-there/

Baharudin, H. (2020, April 9). Coronavirus: No more Zoom for home-based learning after hackers show obscene photos to Singapore students. *The Straits Times.* https://www.straitstimes.com/singapore/hackers-hijack-home-based-lessons-on-zoom-to-allegedly-show-obscene-photos-to-children

Bissell, K., LaSalle, R. M., & Dal Cin, P. (2019). *The cost of cybercrime—Ninth annual cost of cybercrime study*. Technical report, Accenture, 2019. Independently conducted by Ponemon Institute LLC and jointly developed byAccenture.https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-ofcybercrime-study-final.pdf

Cisco. (2018, November 19). Cisco Global Cloud Index: Forecast and Methodology, *2016–2021 White Paper.* Cisco.com. https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1908858

Cyber Security Agency Singapore [CSA]. (2019, September 14). *Third national cybersecurity awareness campaign "Go Safe Online 2019" to boost adoption of good cybersecurity habits*. https://www.csa.gov.sg/news/press-releases/csa-launches-third-national-cybersecurity-awareness-campaign

Field, M. (2018, October 11). WannaCry cyber attack cost the NHS £92m as 19,000 appointments cancelled. *The Telegraph.* https://www.telegraph.co.uk/technology/2018/10/11/wannacry-cyber-attack-cost-nhs-92m-19000-appointments-cancelled/

Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings 1972–1977.* (C. Gordon, L. Marshall, J. Mepham, & K. Soper, Trans.) Pantheon Books. (Original work published 1972–1977).

Fruhlinger, J. (2018, August 30). What is WannaCry ransomware, how does it infect, and who was responsible? *CSO.* https://www.csoonline.com/article/3227906/what-is-wannacry-ransomware-how-does-it-infect-and-who-was-responsible.html

Hootsuite. (2019). *Singapore report: Digital 2019: Singapore* [Slides]. Hootsuite.com. https://hootsuite.com/pages/digital-in-2019

IBM Security. (2019). *The cost of a data breach report*. Independently conducted by Ponemon Institute LLC and jointly developed by IBM Security. https://www.ibm.com/downloads/cas/ZBZLY7KL?_ga=2.118372362.2094645392.15

Koh, D. (2018, March 21). Keynote address by Mr David Koh, Chief Executive, Cyber Security Agency of Singapore, at the 3rd Annual Billington International Cybersecurity Summit. CSA.gov.sg. https://www.csa.gov.sg/news/speeches/the-3rd-annual-billington-international-cybersecurity-summit-keynote-address-by-ce

Kwang, K. (2019, March 6). Cyber Espionage group Whitefly behind SingHealth attack: Symantec. *CNA*. https://www.channelnewsasia.com/news/singapore/singhealth-hack-whitefly-cyber-espionage-group-symantec-11317330

Lapowsky, I. (2019, March 17). How Cambridge Analytica Sparked the Great Privacy Awakening. *Wired.* https://www.wired.com/story/cambridge-analytica-facebook-privacy-awakening/

Morgan, S. (2019). 2019 Official Annual Cybercrime Report: Cybercriminal activity is one of the biggest challenges that humanity will face in the next two decades. *Cybersecurity Ventures sponsored by Herjavec Group*. https://www.herjavecgroup.com/resources/2019-official-annual-cybercrime-report/

Neo, L. S. (2020). Leveraging on digital footprints to identify potential security threats: Insights from the behavioural sciences perspective. In M. Khosrow-Pour (Ed.), *Encyclopaedia of criminal activities and the deep web* (pp. 1,003–1,017). IGI Global. https://doi.org/10.4018/978-1-5225-9715-5.ch068

Ranger, S. (2020, February 3). What is the IoT? Everything you need to know about the Internet of Things right now. *ZD Net*. https://www.zdnet.com/article/what-is-the-internet-of-things-everything-you-need-to-know-about-the-iot-right-now/

Singapore Police Force. (2020, February 5). *Annual Crime Brief 2019*. https://www.police.gov.sg/media-room/statistics

Singapore's cybersecurity startup map. (2020, January 3). *Fintechnews Singapore.* https://fintechnews.sg/34806/security/singapores-cybersecurity-startup-map/

Top 5 biggest cyber attacks in Singapore. (2017, September 29). *Apvera.* https://www.apvera.com/2017/09/29/top-5-biggest-cyber-attacks-in-singapore/

Trend Micro. (n.d.). Cybercriminals. https://www.trendmicro.com/vinfo/us/security/definition/cybercriminals

Vishwanath, A., Neo, L. S., Goh, P., Lee, S., Khader, M., Ong, G., & Chin, J. (2020). Cyber hygiene: The concept, its measure, and its initial tests. *Decision Support Systems*, *128*, 113160. https://doi.org/10.1016/j.dss.2019.113160

Wong, C. (2020, March 3). New inter-ministry committee to fight rise in online scams. *The Straits Times.* https://www.straitstimes.com/politics/singapolitics/new-inter-ministry-committee-to-fight-rise-in-online-scams

This page intentionally left blank

# Part 2

# Cyber Crimes and Cyber Enabled Crimes—Introduction to Emerging Issues

This page intentionally left blank

# Section A

# Violent and Deviant Behaviours Online

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0001

# Chapter 1

# Influence of Social Media on Deviant Acts: A Closer Examination of Live-Streamed Crimes

Whistine Chai Xiau Ting* and John Yu†

*Home Team Behavioural Sciences Centre, Ministry of Home Affairs, Singapore*

**Chai_xiau_ting@mha.gov.sg*

†*john_yu_from.tp@mha.gov.sg*

"The crime scene of tomorrow is going to be the internet of things"

—Mark Stokes

## 1.1 Introduction

On 8 February 2020, Thai soldier Jakrapanth Thomma shot dead 29 people in a mall in north-eastern Thailand and live-streamed the slaughter on Facebook. The 17-hour shooting spree also left more than 50 injured. Throughout the rampage, the gunman was posting updates, photos and videos of his shooting to Facebook's live-streaming platform for close to five hours. During the attacks, he posted statements on Facebook like "No one can escape death" and later asked, "Should I give up?" Facebook

reportedly removed the gunman's profile and posts midway through the crisis after the social media company was alerted by the Thai authorities [News Corp Australia Network, 2020].

Approximately a year before this incident, a similar mass shooting in Christchurch, New Zealand, which resulted in 51 deaths and 50 people injured was live-streamed on Facebook [Saldivia, 2019; Withers, 2019]. Although more than 200 users watched the live-streamed video of the massacre, the illicit content was only alerted to moderators 12 minutes after the broadcast had ended [Sonderby, 2019]. Despite Facebook Live's swift removal of the video after it was flagged, the video of the shooting had gone viral, clocking up more than one million views within the first day. These gunmen's decision to broadcast and display their acts in real-time on the live-streaming platform was unprecedented. After all, filming and live-streaming offending behaviours appear to be counterintuitive and lack logic, as such acts leave a trace for subsequent investigation and arrestment [Sandberg & Ugelvik, 2017].

Undoubtedly, the adage that technology is both a boon and a bane for society and individuals holds true for live-streaming technology. With the release of free mobile streaming video technologies such as Meerkat, Periscope, and Facebook Live in recent years, there has been a surge of internet users streaming or watching digital video content online [Statista, 2018]. These platforms offer users the convenience of distributing and delivering content in real time. The ability of anyone to broadcast his or her activities to global audiences imprudently has also spilled over to the domain of criminal offences. From 2015 to 2017, there were at least 45 live-streamed instances of violence reported on Facebook Live alone [Kantrowitz, 2017]. This live broadcast of criminal activities over social media is now known as live-streamed crime. This chapter will explore the influence of social media on deviant acts, with a focus on the motivations underlying live-streamed crimes. Various means to mitigate and combat live-streamed crimes will also be proposed.

## 1.2 Live-streamed Crime may Involve Performance Crime

A live-stream refers to any video footage that is both recorded and broadcasted simultaneously through digital media [Payne *et al*., 2017]. Live-streamed crime is therefore any criminal behaviour that is recorded

and broadcasted simultaneously via live-stream technology. Live-streamed crimes represent a new face of crime in the modern era as criminal acts conducted in the real world are being witnessed in cyberspace in real-time. The content might include pre-crime footage of a criminal leaking his intentions, actual footage of the crime as it happens, or post-crime footage of confessions or bragging [Yu *et al*., 2020] Since live-streaming services are now widely utilised by individuals, communities, and industries for various purposes (whether to gain popularity, gather like-minded individuals or to promote one's agenda), it should not be surprising that incorporating live-stream into criminal activities shares similar motivations, which will be explored in-depth in this chapter.

Sometimes, live-streamed crime is also referred to as performance crime. Performance crime is any recording, sharing, and uploading of crime with the purpose of distributing the "act" to new media audiences [Surette, 2015; Yar, 2012]. Surette [2015] suggested that the rise of performance crime could be linked to the pervasiveness of modern celebrity culture, where one aspires to achieve celebrity-like status. These criminal acts are committed with an audience in mind and filmed with the intention of achieving certain goals or functions such as self-promotion or causing fear to communities. On the other hand, some actors of the performance crime may be unwittingly involved in the production without their knowledge or consent [Surette, 2015]. Performance crimes are often self-incriminatory and contain details that can be utilised in the identification, apprehension, and prosecution of the perpetrators. An example is the video recordings of acts of violent extremism that are intentionally distributed to news outlets or online sources with a statement by the actor or group claiming responsibility and justifying their actions [Yu *et al*., 2020].

Figure 1.1 highlights the differences between performance crime and live-streamed crime. One may argue that they are two sides of the same coin, since both involves recording and broadcasting of the crime. However, not all performance crimes are live-streamed, and not all live-streamed crimes involve a performative element. An example of the latter would be the voyeuristic live-stream recording of victims in private places without a performative element [Yu *et al*., 2020]. These perpetrators do not seek to self-promote through public dissemination but do so underhandedly within closed communities with other goals such as self-gratification or profit [Jeong & Griffiths, 2019]. In this chapter, the focus will be on live-streamed crime.

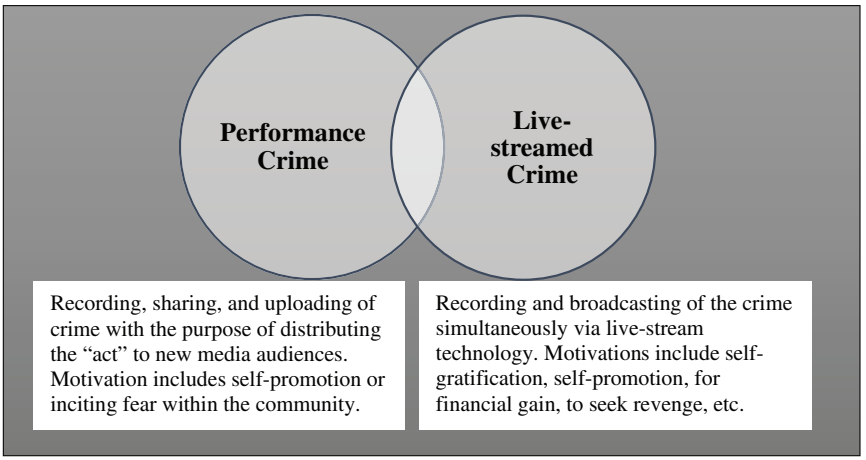



Figure 1.1. Differences between performance crime and live-streamed crime.

## 1.3 The Diverse Nature of Live-streamed Crimes

From the 2019 Christchurch shooting to the recent Thai mall shooting, live-streamed crimes appear to be the inevitable consequence of rapid technological change, as offenders incorporate new technologies into their modus operandi. In an attempt to gain better insights of the types of crime that utilise live-streaming technology, Yu, Chai, and Tambagan [2020] reviewed a total of 33 crime cases reported by online news media from January 2015 to July 2019. A horizon scan was conducted to sieve out crimes or law-breaking incidents involving a live-streaming component or platform prior to, during, and after the crime. In their review, the various live-streamed crimes were further categorised based on their functions. The study revealed that the use of live-streaming platforms to commit crime is extensive and diverse. Startlingly, there were as many as 15 different types of crimes involving live-streaming elements. These included online sexual exploitation (21.2%), murder or grievous hurt resulting in death (12.1%), digital piracy (9.1%), rape (9.1%), acts of violent extremism (6.1%), animal abuse (6.1%), assault (6.1%), child physical abuse (6.1%), murder-suicide (6.1 %), driving under influence of alcohol (3.0%), drug use (3.0%), exhibitionism (3.0%), housebreaking (3.0%), "swatting" resulting in death (3.0%), and voyeurism and distribution of obscene materials (3.0%) [Yu *et al*., 2020].

As highlighted by the 33 crime cases, live-streaming can occur within a diverse range of offending behaviours. Online sexual exploitation was most frequently cited, followed by murder or grievous hurt resulting in death, rape, and digital piracy [Yu *et al*., 2020]. Likewise, the underlying motivations behind the offenders' decision to live-stream their deviant acts are also diverse. The purpose could be instrumental (e.g., for financial gains) or expressive (e.g., for self-gratification or self-promotion). The next section will discuss the reasons why offenders broadcast their acts.

## 1.4 Perpetrators Turned to Live-streamed Crimes for Diverse Reasons

With live-streamed technology, it is now easier, cheaper, and almost instantaneous for perpetrators to produce and distribute digital content. In view of these advantages, many perpetrators have now turned to live-streaming platforms. Figure 1.2 outlines the various reasons underlying live-streamed crimes.

### 1.4.1 *Lucrative modus operandi for cyber-facilitated criminality*

With live-stream technology, it is now easier, cheaper, and almost instantaneous for perpetrators to produce and distribute digital content.

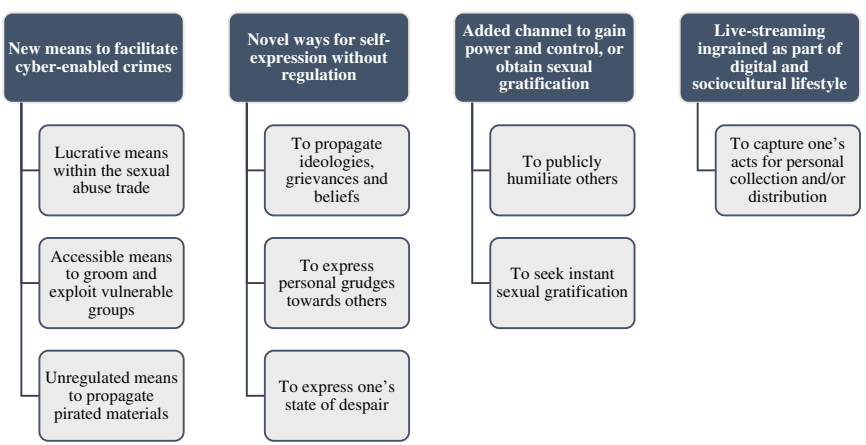



Figure 1.2. Various reasons behind the live-streaming of crimes or deviant behaviours.

In view of these advantages, many perpetrators have now turned to live-streaming platforms. Since live-streaming represents a new leap in technology that enables simultaneous production and distribution of digital content, it was only a matter of time before crimes that rely on unlawful recording and broadcasting began leveraging on this advancement.

A closer look at the online sexual abuse trade in the Philippines gives us further insight into how live-streaming has become a lucrative means for this criminal activity [Promchertchoo, 2018]. With one in five Filipino children falling prey to online sexual exploitation, the Philippines has been dubbed as the "global epicentre of the live-stream sexual abuse trade" [Brown, 2016]. In particular, localities with high poverty rates and strong internet culture have become operating grounds for the child sex tourism industry. Fuelled by poverty, some parents or family members may even encourage or compel their children to participate in this online sex trade operated by syndicates. Cyber-sex predators from all over the world would pay and direct the victims, through online platforms, to perform sexual acts with other juveniles—such as a sibling—or adults. These paedophiles may then access such content illicitly through the dark web, where they can stream these contents, and at times even direct the course of the sexual abuse based on the amount they are willing to pay [Scotland & Glasgow, 2019]. In addition, as live-streams do not need to be downloaded and stored on devices, there is no evidence stored on the viewers' devices, giving offenders an additional safeguard against criminal prosecution [Yu *et al*., 2020].

At the same time, online sex predators groom and exploit vulnerable groups to perform explicit acts for their own gratification. Children and teenagers who use live-streaming platforms on a widespread and frequent basis are the most vulnerable targets. The grooming and exploitation may occur within social media outlets, such as Bigo Live and Periscope, where predators operating alone or as a group would search for opportunities to groom children remotely and anonymously across the internet. These acts often go undetected or unreported if the viewers do not report the inappropriate contents to the relevant authorities [Yu *et al*., 2020]. Furthermore, these perpetrators may continue to persuade the more vulnerable victims to create private channels via other platforms such as Skype where they can further exploit them with greater anonymity and privacy [Bevan, 2019; Smith, 2019; Wheatstone, 2019].

Live-streaming technology also offers a new platform for operators and consumers of digital piracy to provide and access live entertainment for free. Inevitably, such technology has been abused to propagate the

illegal filming and distribution of copyrighted content. A recent example in 2019 is the rampant live-streaming of episodes of the popular television series, Game of Thrones, in the "Just Chatting" section of Twitch, a popular streaming platform for gamers. It was reported that one streamer openly live-streamed the show to more than 500 audience members [Alexander, 2019; Yu *et al*., 2020]. Interestingly, some sub-communities on Twitch endorsed these pirated streams and even provided strategies to streamers on how they could avoid detection while doing so [Yu *et al*., 2020].

### 1.4.2 *Accessible means of self-expression without regulation*

A decade ago, Powell [2010] described how the internet has become a significant tool for self-presentation as we create personalities and images of our selves through the lens of social media. Now, with the advent of live-stream technology, the internet offers an even easier avenue for disseminating unregulated and unbounded content for the purpose of self-expression.

With the profusion of live-streaming applications and capabilities online that allow one to freely and easily create and transmit content, anyone can have direct access to an extensive global audience and deviant subcultures. Live-streamed platforms provide offenders and deviant groups with the capacity to broadcast their ideologies, grievances, or illicit materials freely to the masses. Such platforms also allow anyone to gain notoriety without regulation. Many groups of violent extremists have exploited such technological advances to broadcast their ideologies, hate speeches, or videos to promote and gather a large pool of followers within a short span of time [Bender, 2019]. Notably, while much attention is focused on live-streamed violence by violent extremists, this group makes up only a small proportion of live-streamed crimes compared to the other acts of violence committed by individuals or small groups acting out of their own personal beliefs and attitudes [Yu *et al*., 2020].

Evidently, live-streaming has become a portable method for violent individuals to express their thoughts and emotions while committing their crimes. In 2017, Wuttisan Wongtalay, a depraved father who suspected his wife of having affairs, hanged their 11-month-old daughter live on Facebook before hanging himself. This murder–suicide in Phuket was an act of jealousy, and the graphic footage remained on Wuttisan's Facebook page for almost a day before it was removed. Although the original video was removed, the appalling screenshots had circulated widely and

continue to remain on the social media platform [Hodge, 2017]. The shocking footage emerged just days after the infamous live-streamed murder of Robert Godwin by Steve Stephens, and both incidents are just one of the many live-streamed murders that had gone viral [Selk, 2017]. In 2017, Steve Stephens had randomly shot Robert Godwin before shifting the blame to his lover, Joy Jane. In both cases, the perpetrators live-streamed their acts in an attempt to aggravate their loved ones and inflict guilt on them. The use of live-streaming has thus become a means for any individual to seek revenge or express any pent-up emotions in front of a readily available audience via online platforms [Yu *et al*., 2020].

The ability to express oneself freely to a responsive audience may also explain the use of live-streaming prior to the act of suicide. In Turkey, Ayhan Uzun recorded himself on Facebook Live expressing his grievance and disappointment that his daughter had married without his consent. He then pointed a handgun to his head and threatened to commit suicide. In the end, Uzun ended his life despite his family and friends' frantic dissuasion during the live-stream [Robinson, 2017]. With live-streams, suicide no longer becomes a private affair but a disturbing spectacle for family, friends, and even strangers. The live-streaming of suicides may function as a 'cry for help' from available digital sympathisers as the final barrier to suicide. For some, this live-stream may also be a modern substitute for a suicide note addressed to specific individuals or to a larger public to justify the act [Yu *et al*., 2020].

### 1.4.3 *An added channel to gain control or obtain sexual gratification*

Next, live-stream platforms can offer another avenue for individuals to openly humiliate others. Perpetrators may feel a sense of power and control when they victimise their subjects particularly in violent crimes such as sexual assault. The use of live-streaming exacerbates victim-shaming as the victims' identities are not concealed and they are publicly humiliated and dehumanised before a live audience. The 2017 Uppsala rape in Sweden is a prominent example, where a group of three men raped an unconscious woman over several hours and live-streamed the act to a closed group of 60,000 members on Facebook. The live-stream was only intercepted when the police arrived at the scene after viewers had reported the crime. During the live-streaming of the rape, approximately 200 people out of the 60,000 members were online at the time of the incident [Kale, 2017; Scally, 2017]. To make matters worse, live-streams of such

crimes may also be recorded and shared numerous times after the live viewing of the event. While the original posts have been taken down, copies may be kept and posted on other websites to be re-watched by many others, which will exacerbate the secondary trauma and humiliation experienced by the victims [Yu *et al*., 2020].

It is possible that apart from exerting control and power, some perpetrators intentionally live-stream the sexual acts in order to obtain sexual gratification. With the rise of dating websites, sexual social networks, and online pornography, there has been a shift of video sharing of sexual contents from a unidirectional "broadcasting mode" by producers to an interactive "social media mode" by individuals [Kreps, 2010]. This social media mode implies that individuals intentionally create their own sexual materials, sometimes without the consent of their partners, for others to view and comment on. Kreps [2010] further argued that the fundamental dyad of exhibitionism and voyeurism could be present in such phenomenon. Indeed, live-streaming applications such as Meerkat and Periscope have unveiled the new norms of exhibitionism and voyeurism, where private sexual acts or even group orgies were live-streamed for both the actors and viewers' instant gratification. One can expect the recording and broadcasting of online pornography or sexualised violence to further proliferate in the future.

### 1.4.4 *Live-streaming usage as part of one's digital and sociocultural lifestyle*

Finally, the use of live-streams may simply reflect the lifestyle of a contemporary, digitally connected generation, including criminals. It is fast becoming a social norm for people from all walks of life to share their views and aspects of their lives online, so much so that engagement in social media has turned into an integral part of social interaction and identity formation for many [Lai *et al*., 2020]. One example is the emergence of lifestyle streamers, who live-stream a broad variety of interests, from cooking to travel. [Lai *et al*., 2020]. As these streamers become producers and distributors of their own content, some may create postings that violate legal or moral norms of society knowingly or unknowingly. It is likely that criminals who live-stream their crimes represent a small, deviant subgroup who overshare their criminal activities online for social validation even if it backfires on them when the footage is released and incriminates them [Yu *et al*., 2020].

According to Sandberg and Ugelvik [2017], sociocultural explanations may shed some light on why criminals would record their crimes. Firstly, with the ubiquity of smartphones and other portable capturing devices, the practice of instant photo-taking and sharing on digital platforms is becoming a mainstay in modern society. In particular, recording events or behaviours that are out of the ordinary and elicit surprise or shock has become second nature for many. Crime would likely fall within this category as an exceptional event that elicits strong emotions, and criminals might be motivated by the same urge to capture the situation for personal collection or distribution. Secondly, the increasing sexualisation of digital content due to the prevalence of multiple genres of sexual and sexually-explicit materials in the web can distort one's understanding of what is accepted [Davis, 2018]. Perpetrators might have been influenced by a perceived mainstreaming of online pornography or sexualised violence in modern society and use it to justify their use of live-streaming services to grow their repository with their personal recordings.

## 1.5 Live-Streamed Crime Illustrates How Social Media has Transformed Crime and Crime Investigation

With the advent of social media, the way an individual offends in the eyes of the audience has transformed drastically. The use of mobile streaming video technologies has also blurred the boundary between public and private space. As video production and distribution have become almost simultaneous, new challenges in terms of policing and preventing disturbing broadcasts from being circulated have emerged. In an era dominated by social media, the ability of crimes to exist both offline and online, as well as the potential for disturbing footages of criminal acts to circulate uncontrollably have important implications for law enforcement agencies around the globe.

### 1.5.1 *Digital evidence of live-streamed crime is fluid, multi-directional, and owned by many*

First, live-streamed crime implies that the official ownership and control of crime-related media content by the perpetrators is lost. Traditionally,

evidence of crime tends to be textual, paper-based, and linear in nature. Now, live-streamed crimes leave behind evidence that is digital, multimedia, and multidirectional; the crime is not confined to any physical crime scene, and can be recorded live, witnessed by many simultaneously, and circulated virally across continents. While the act might have been streamed for a small target audience (e.g., within online sexual exploitation groups), access is often limitless because of its digital nature [Strutin, 2011; Surette, 2015]. The "crime scene/evidence" has become more fluid and possessed by many. This poses operational and investigative challenges for law enforcement officers as the owners of the content and perpetrators of the crime may no longer be that apparent [Strutin, 2011; Surette, 2015; Yu *et al*., 2020].

### 1.5.2 *Viewers have transitioned from passive audiences to active abettors of the crime*

Second, we witness a shift in the level of influence and involvement of the online public in the act of the live-streamed crime. The isolated act of watching the news through traditional media such as newspaper and television has been replaced by the joint experience of "posting, tweeting and going rival" online [Surette, 2015]. Content consumers can now be the producers and distributors of self-generated content [Surette, 2015]. Netizens are now not only passive witnesses of a criminal act, but can also easily become "influencers" through posting feedback and influencing the course of the crime immediately, or "distributors" by sharing, tweeting or circulating the contents to others [Surette, 2012]. Such activities may be perceived to be endorsing the crime and further inciting the perpetrator to intensify the severity and frequency of the act. With this element of audience participation that results in social reinforcement, those who live-stream their crimes to pursue notoriety and attain online celebrity status would be further encouraged to continue their deviant acts.

### 1.5.3 *Delay between production and distribution of live-streamed contents is almost non-existent*

While live-streamed crimes have paved new ways for law enforcement agencies to prevent and combat crimes through cyber threat assessment and tracing of digital evidence, such crimes also pose challenges to their operations. Specifically, the reduction in duration between the time when

a crime is committed and reported after being witnessed on a live-stream may influence the public's attitudes towards law enforcement and perceptions of their efficacy. Immediate post-crime reactions of fear or distress, particularly for serious crimes, may result in an outcry against the lack of swiftness in intervention by law enforcement agencies [Yu *et al*., 2020]. For example, when the 2020 Thai mall shooting rampage occurred, some of the victims pored through social media and made frantic calls to the authorities while trapped inside the mall. Furthermore, news outlets were providing live coverage of the attack, which might not only have jeopardised tactical operations, but also mounted greater pressure on the police to swiftly intervene and arrest the gunman ["Thailand shooting: How the massacre unfolded," 2020]. Finally, tactics and processes utilised by law enforcement that are captured on live-streams may be subjected to biased scrutiny by uninformed live commentators.

## 1.6 There are Multiple Ways We Can Respond to, Combat, and Prevent Live-streamed Crimes

Undeniably, social media has presented both challenges and opportunities to law enforcement, and agencies can utilise information that is publicly accessible on social media to combat and prevent live-streamed crimes. This section explores what has already been done and proposes some possible ways for enforcement agencies to mitigate or respond more effectively to live-streamed crimes.

### 1.6.1 *Law enforcement to enhance efforts on cyber surveillance and risk management*

Foremost, cyber surveillance and risk management should be one of the key focus of law enforcement agencies in combating live-streamed crimes. As threats can surface rapidly, law enforcement agencies should continue to hone their capabilities in risk detection, reduction, and pre-emptive interventions [Surette, 2015]. Anyone operating within the cyberspace inevitably leaves behind publicly available digital footprints or evidence, which can be used in cyber threat assessments conducted by intelligence and law enforcement agencies [Whitty *et al*., 2017]. Given the ubiquity of digital evidence (e.g., photographs and videos of crime,

transfer of funds online, publicised social networks, online expression of ideology and intent, hate messages on social media), law enforcement can harness this vast information available for surveillance and risk assessment.

### 1.6.2 *Law enforcement to collaborate with digital platforms and firms*

For the same reason, law enforcement should work closely with technology firms and social media platforms with live-streaming capabilities to fine-tune methods of recognising, intervening, and preventing the live-streaming of offending behaviours by the perpetrators [Yu *et al*., 2020]. This may involve proactive surveillance of the online activities of persons or forums of interest known to incubate discussions or circulate materials that are disconcerting or inimical to societal values. Collaboration with technology firms is also vital to expand the watch list for individuals or groups of interest, to better understand the networks of those disseminating illicit content in cyberspace, and to be kept abreast of new criminal behaviours or trends that surface online. To achieve this, law enforcement can work with local branches of tech firms to set up processes to facilitate the flow of information pertinent to public safety and security, such as encouraging firms to report suspicious users or groups who have been repeatedly flagged for violations of relevant community standards. Further legislative measures might be necessary for firms to comply with such investigations.

In the aftermath of the 2019 Christchurch shooting, additional measures were implemented by Facebook such as tightening of guidelines, addressing warning signs, and restricting usage by those who violate rules and regulations [Abbruzzese, 2019; Hermesauto, 2019]. However, after the 2020 Thai mall shooting incident, the social media giant again received criticism for the lack of a swift response in blocking and removing the footage of the massacre [Afp, 2020]. As such, it is critical for law enforcement to work closely with technology firms and social media companies, understand their challenges and limitations, and to persuade them to build up their capabilities to support online surveillance and risk assessment. Other safeguards may include expedited processes to report, regulate, and remove disturbing content upon identification, as well as technological measures to restrict or prohibit online participation

by netizens in order to reduce the incentive for perpetrators to live-stream their criminal activities.

To encourage greater partnership, engagement, and information sharing between social media companies and law enforcement agencies, regular forums on best practices, challenges, limitations, and strategies of cyber risk assessment and management would be useful. Close collaboration between multiple agencies will enable a quicker response to any potential live-streamed crimes and greater knowledge of details of likely attacks. Such collaboration provides shared situational awareness of any impending threats, which will result in faster and more coordinated assessments and responses to the threat [Johnson *et al*., 2016; Levi *et al*., 2017].

### 1.6.3 *Government to enhance efforts to raise public awareness, vigilance, and reporting amongst the online community*

In addition, involving the community as part of a concerted effort against the spread of live-streamed crimes is pertinent. In the event that surveillance and technological means have failed to detect the signs or the actual live-stream, the online community would be the first to observe the incident. Thus, it is crucial that users report offences or problematic content to the social media or law enforcement agency when they come across such content online. It will be useful to allocate more funds and resources to raise awareness of the role of the online community in combating live-streamed crime [Yu *et al*., 2020]. For example, schools can raise greater awareness of the impact of social media on deviant acts and how students should respond to them. This, among other means, would encourage greater reporting of the production and circulation of the contents of live-streamed crime. Encouraging greater reporting also reduces any online bystander effect that could occur in the event of a crime being live-streamed [Hendricks, 2014; Hudson & Bruckman, 2004].

### 1.6.4 *Researchers and academia to conduct further research to gain greater insights on live-streamed crime*

The rise of live-streamed crimes also calls for more extensive research on this phenomenon. As research into and knowledge about live-streaming

crimes remain limited, future research can explore in greater detail the thought processes and motivations underlying deviant behaviours that are live-streamed. For now, research on live-streamed crime is still in its infancy, but it is expected that greater insights will be generated as studies of new cases of live-streamed behaviours emerge. Since live-streamed crimes occur across many categories of crime types, future research could dwell into which deterrence and preventive factors and measures might be useful. Another important angle of research could include the profile of victims and the impact of live-streamed footages on them, which will better inform the type of support and recovery services that should be offered to them. More research grants by both the government and tech firms could be provided for researchers to conduct more studies on live-streamed crimes.

### 1.6.5 *Law enforcement to introduce more targeted crime prevention efforts*

Finally, crime prevention efforts should also target crimes that utilise live-streaming platforms for various gains. For instance, given that many live-streamed crimes committed to date are for self-gratification or self-promotion purposes, crime prevention and management efforts could be targeted towards crimes underlying such motivations. One good example is the handling of the Christchurch shooting by the New Zealand government in 2019. Given that the perpetrator hoped to gain notoriety from the shooting, New Zealand discouraged the country from fulfilling the gunman's wishes by removing the perpetrator's names, photos, and identities from all media publicity, thereby denying him of the notoriety he sought after [Merelli, 2019]. In addition, based on the study by Yu, *et al*. [2020], the majority of live-streamed crimes reported are those that sexually exploit victims online. With the high social media penetration rate among children and teenagers in most developed countries, such as Singapore [Grosse, 2018], there is a greater need for closer monitoring of online child sexual exploitation and grooming occurring on popular digital platforms. Combating such crimes through preventive efforts such as public awareness or education may help to reduce the rate of victimisation among vulnerable individuals. Crime prevention efforts introduced should be informed by more extensive research on live-streamed crimes.

## 1.7 Conclusion

This chapter takes a closer look at the nature and phenomenon of live-streamed crimes that has surfaced in recent times. Based on existing literature, live-streaming technology has been applied across a spectrum of criminal activities. This is due to the convergence of the ubiquity of mobile-streaming applications, the increasing connectedness that modern, deviant individuals enjoy with communities in cyberspace, as well as the fulfilment of intrinsic and extrinsic motivations that simultaneous live-streams can offer to the producer or consumer of such technology.

The union of live-streaming and crime is the product of the modern era and would thereby require contemporary or forward-looking measures to address the many harms that it can cause to individuals and society. Given the spontaneity at which live-streamed crimes can appear in cyberspace, and the vast extent of its reach across transnational boundaries, it is crucial for a comprehensive partnership between intelligence and enforcement agencies, technology companies, and online communities to be established in order to better contain and control live-streamed crimes. This can also be better advised with further research into live-streaming behaviours and outcomes that can aid in the development of better interventions and preventative measures.

## 1.8 Acknowledgement

The views expressed in this chapter are the authors' only and do not represent the official position or view of the Ministry of Home Affairs.

## 1.9 References

Abbruzzese, J. (2019, March 19). In New Zealand shooting aftermath, tech's role in spreading extremism comes under scrutiny. *NBC News*. https://www.nbcnews.com/tech/tech-news/new-zealand-shooting-aftermath-tech-s-role-spreading-extremism-comes-n984336

Afp. (2020, February 10). Thai mall gunman posted to Facebook 'for hours' as deadly rampage continued. *News.com.au*. https://www.news.com.au/world/asia/thai-soldier-kills-many-in-shooting-rampage/news-story/15150801a4bfbe95da90ad921b6e1a0b

Agence France-Presse. (2016, May 11). Periscope used by French teenager to live-stream her own suicide. *The Guardian*. https://www.theguardian.com/

world/2016/may/11/french-prosecutors-investigate-woman-live-streams-suicide-twitter-periscope

Agence France-Presse. (2018, December 28). Six Men Arrested in Vietnam for Killing and Eating Endangered Monkey on Facebook Live-stream. *South China Morning Post*. https://www.scmp.com/news/asia/southeast-asia/article/2179881/six-men-arrested-vietnam-killing-and-eating-endangered

Alexander, J. (2019, April 30). People are Live-Streaming New Game of Thrones Episodes on Twitch Every Week. *The Verge*. https://www.theverge.com/2019/4/30/18485658/game-of-thrones-season-8-twitch-stream-hbo-now

Ali, S. S. (2017, March 22). Gang Rape of Chicago Teen Was Watched Live by 40 People on Facebook, No One Called Cops. *NBC News*. https://www.nbcnews.com/news/us-news/gang-sex-assault-chicago-teen-was-watched-live-40-people-n736616

Bender, S. (2019, March 20). Commentary: Social media and live-streaming have created performance terrorism. *CNA.* https://www.channelnewsasia.com/news/commentary/christchurch-attack-shootings-social-media-video-live-streaming-11362152

Bevan, A. (2019, June 20). Child Sexual Abuse is Being Live-streamed Right Now to a Paedophile Near You. *The Scotsman*. https://www.scotsman.com/news/crime/andrew-bevan-child-sexual-abuse-is-being-live-streamed-right-now-to-a-paedophile-near-you-1-4950494

Brown, A. (2016, October 19). Safe from harm: Tackling online child sexual abuse in the Philippines. *Unicef*. https://www.unicef.org/protection/philippines_91214.html

Cancian, D. (2019, 17 May). Florida man lives-streamed argument with Walgreens staff before smacking baby in head: 'I didn't do nothing!'. *Newsweek*. https://www.newsweek.com/ryan-greenlee-facebook-livestream-crime-largo-police-department-1428579

CBC News. (2017, April 12). Saskatoon man charged in international child porn ring faces new charges. *CBC News*. https://www.cbc.ca/news/canada/saskatoon/philip-chicoine-new-child-porn-charges-1.4067509

CBS San Francisco. (2019, January 20). Burglar Live Streams Own Crime On Victim's Phone, Strips To Underwear. *CBS SF*. https://sanfrancisco.cbslocal.com/2019/01/20/burglar-live-streams-own-crime-on-victims-phone-strips-to-underwear/

Cheng, K. (2018, December 20). Chinese man live-streams himself torturing small dogs with a TASER to earn money from followers before being caught and publicly shamed. *Mail Online*. https://www.dailymail.co.uk/news/article-6516531/Chinese-man-live-streams-torturing-small-dogs-TASER-caught-shamed.html

Criss, D. (2017, April 18). The Facebook victim was a granddad walking home after Easter meal. *CNN*. https://edition.cnn.com/2017/04/17/us/facebook-homicide-victim-trnd/index.html

Davis, S. E. (2018). Objectification, Sexualization, and Misrepresentation: Social Media and the College Experience. *Social Media + Society, 4*(3). https://doi.org/10.1177/2056305118786727

England, C. (2017, February 15). Teenager jailed for broadcast of girl's rape on online Periscope app. *The Independent*. https://www.independent.co.uk/news/world/americas/teenager-marina-lonina-livestream-rape-17-year-old-friend-periscope-app-sentence-prison-columbus-a7581196.html

Grosse, S. (2018, October 19). Nudity, public sex, stalkers: What children are in for on live-streaming apps. *CNA*. https://www.channelnewsasia.com/news/cnainsider/parents-children-live-streaming-apps-online-porn-virtual-gifts-10838726

Hendricks, V. F. (2014, June 5). The 21st century bystander effect happens every day online. *The Conversation*. https://theconversation.com/the-21st-century-bystander-effect-happens-every-day-online-27496

Hermesauto. (2019, May 15). Facebook restricts live-streaming feature, citing New Zealand shooting. *The Straits Times*. https://www.straitstimes.com/world/united-states/facebook-restricts-live-feature-citing-new-zealand-shooting

Hodge, M. (2017, April 25). Evil knows no bounds. Depraved dad hangs himself and his 11-month-old baby on Facebook Live after accusing his girlfriend of cheating. *The Sun.* https://www.thesun.co.uk/news/3407726/dad-hangs-baby-facebook-live-phuket-thailand/

Hodge, M. (2017, January 29). Sick sexual predators are grooming young children on livestreaming app Periscope. *The Sun*. https://www.thesun.co.uk/news/2734074/paedophiles-little-girl-live-streaming-periscope/

Hudson, J. M. & Bruckman, A. S. (2004). The bystander effect: A lens for understanding patterns of participation. *The Journal of the Learning Sciences*, *13*(2), pp. 165–195.

Jeong, S. & Griffiths, J. (2019, March 21). Hundreds of motel guests were secretly filmed and live-streamed online. *CNN*. https://edition.cnn.com/2019/03/20/asia/south-korea-hotel-spy-cam-intl/index.html

Johnson, C., Badger, L., Waltermire, D., Snyder, J., & Skorupka, C. (2016). *Guide to cyber threat information sharing*. National Institute of Standards and Technology. http://dx/doi.org/10.6028/NIST.SP.800-150

Kale, S. (2017, January 23). Three Men Arrested on Suspicion of Live-Streaming Gang Rape on Facebook. *Vice*. https://www.vice.com/en_us/article/j5epd8/three-men-arrested-suspicion-live-streaming-gang-rape-facebook

Kantrowitz, A. (2017, June 16). Violence on Facebook Live is Worse than You Thought. *Buzzfeed News*. https://www.buzzfeednews.com/article/alexkantrowitz/heres-how-bad-facebook-lives-violence-problem-is#.bwo251N9m

Kreps, D. (2010). Foucault, exhibitionism and voyeurism on chatroulette. *In Cultural Attitudes towards Technology and Communication*, 2010, pp. 207–216.

Lai, J., Chen, A., & Murrow, L. (2020, April 17). The Rise of Lifestyle Streamers. https://a16z.com/2020/04/08/lifestyle-streamers/

Levi, M., Doig, A., Gundur, R., Wall, D., & Williams, M. (2017). *Cyberfraud and the implications for effective risk-based responses: themes from UK research*. Crime Law Soc Change. 67, pp. 77–96.

Levin, S. & Jamieson, A. (2017, January 5). Four suspects charged with hate crimes over beating in Facebook Live video. *The Guardian*. https://www.theguardian.com/us-news/2017/jan/05/facebook-live-beating-anti-donald-trump

McCausland, P. (2017, May 15). Memphis, Tennessee, Man Fatally Sets Himself on Fire on Facebook Live. *NBC News*. https://www.nbcnews.com/news/us-news/memphis-tennessee-man-fatally-sets-himself-fire-facebook-live-n759366

Mclaughlin, K. (2017, February 15). Two Slovenian men are arrested after victim dies following sickening 20-minute beating that was streamed on Facebook Live. *Mail Online*. https://www.dailymail.co.uk/news/article-4228048/Slovenian-men-arrested-Facebook-Live-beating.html

McPhate, M. (2016). Teenager Accused of Live Streaming a Friend's Rape on Periscope. *The New York Times*. https://www.nytimes.com/2016/04/19/us/periscope-rape-case-columbus-ohio-video-livestreaming.html

Menendez, E. (2019, March 19). Ex-deputy headteacher admits taking class A drugs while live-streaming child sex abuse. *Metro*. https://metro.co.uk/2019/03/19/ex-headteacher-admits-taking-class-a-drugs-while-live-streaming-child-sex-abuse-8951783/

Merelli, A. (2019, March 19). New Zealand prime minister's brave lesson on how to deny terrorists fame. *Quartz*. https://qz.com/1572719/new-zealand-prime-minister-refuses-to-name-christchurch-shooter/amp/

Michael, T. (2017, May 4). Serena McKay, 19, 'beaten to death by gang of schoolgirls who may have filmed themselves stomping on her head in sick Facebook Live video that remained on the site for FOUR HOURS'. *The Sun*. https://www.thesun.co.uk/news/3478789/serena-mckay-beaten-to-death-facebook-live-video/

Nation Thailand, The. (2019, March 19). Deaf-mute Thai man commits suicide on Facebook Live. *The Nation Thailand*. https://www.nationthailand.com/around_thailand/30366108

Payne, K., Keith, M. J., Schuetzler, R. M., & Giboney, J. S. (2017). Examining the learning effects of live streaming video game instruction over Twitch. *Computers in Human Behavior, 77*, pp. 95–109.

Philips, K. (2017, January 15). A 12-year-old girl live-streamed her suicide. It took two weeks for Facebook to take the video down. *The Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2017/01/15/a-12-year-old-girl-live-streamed-her-suicide-it-took-two-weeks-for-facebook-to-take-the-video-down/

Powell, A. (2010). Configuring Consent: Emerging Technologies, Unauthorized Sexual Images and Sexual Assault, *Australian and New Zealand Journal of Criminology*, *43*, pp. 76–90.

Promchertchoo, P. (2018, October 27). Live-streaming of child sex abuse spreads in the Philippines. *CNA*. https://www.channelnewsasia.com/news/asia/philippines-child-sex-abuse-live-streaming-cybersex-exploitation-10769092

Quigley, C. (2019, March 12). Police: 2 Raleigh men live streamed drug use with child in room. *CBS 17*. https://www.cbs17.com/news/local-news/wake-county-news/police-2-raleigh-men-live-streamed-drug-use-with-child-in-room/

Reuters. (2019, May 15). Facebook restricts Live feature, citing New Zealand shooting. *CNA*. https://www.channelnewsasia.com/news/business/christchurch-attack-facebook-live-ardern-11534738

Roberts, J. J. (2016, June 3). This Happened When a Guy Streamed a Movie to Facebook Live. *Fortune*. https://fortune.com/2016/06/03/facebook-live-movie-streaming/

Robinson, J. (2017, October 24). Turkish father livestreams his suicide on Facebook because his daughter chose to marry without his approval. *Mail Online*. https://www.dailymail.co.uk/news/article-5011883/Father-kills-Facebook-Live-marriage-plan.html

Rubin, A. J. & Blaise, L. (2016, June 14). Killing Twice for ISIS and Saying So Live on Facebook. *The New York Times*. https://www.nytimes.com/2016/06/15/world/europe/france-stabbing-police-magnanville-isis.html

Saldivia, G. (2019, March 16). Number Of Dead Rises To 50 In New Zealand Mass Shooting. *NPR*. https://www.npr.org/2019/03/16/704133810/number-of-dead-rises-to-50-in-new-zealand-mass-shooting

Sampathkumar, M. (2017, December 30). Kansas 'swatting' latest: Man arrested in Los Angeles after gamer Andrew Finch shot dead by police in Wichita. *The Independent*. https://www.independent.co.uk/news/world/americas/innocent-gamer-shot-dead-kansas-wichita-swatting-andrew-finch-a8134521.html

Sandberg, S. & Ugelvik, T. (2016). Why do offenders tape their crimes? Crime and punishment in the age of the selfie. *British Journal of Criminology, 57*(5), pp. 1,023–1,040.

Scally, D. (2017, January 25). Three for court in Sweden over Facebook Live 'gang-rape'. *The Irish Times*. https://www.irishtimes.com/news/world/europe/three-for-court-in-sweden-over-facebook-live-gang-rape-1.2950554

Scotland, W. & Glasgow. (2019, July 2). Man admits live-streaming child sex abuse at Irvine home. *BBC News*. https://www.bbc.com/news/uk-scotland-glasgow-west-48841262

Selk, A. (2017, April 19). The 'Facebook Killer' is Dead—But the Hate Against His Ex-Girlfriend Lives On. *The Washington Post*. https://www.washingtonpost.com/news/post-nation/wp/2017/04/19/the-facebook-killer-is-dead-but-the-hate-against-his-ex-girlfriend-lives-on/?utm_term=.159e5eae07c3

Smith, T. M. (2019, June 20). Appeals Court Upholds 40-Year Sentence for Former Maypearl Police Chief, Child Predator. *Daily Light*. https://www.waxahachietx.com/news/20190620/appeals-court-upholds-40-year-sentence-for-former-maypearl-police-chief-child-predator

Sonderby, C. (2019, March 18). Update on New Zealand. *Facebook Newsroom*. https://newsroom.fb.com/news/2019/03/update-on-new-zealand/

Spangler, T. (2015, April 4). Twitter's Periscope and Meerkat Invade Theatres, But Movie Biz Not Too Worried. *Variety*. https://variety.com/2015/digital/news/twitter-periscope-meerkat-movie-theater-livestream-furious-7-1201466367/

Spargo, C. (2015, October 12). Florida woman arrested after she starts live streaming on Periscope to tell the world she' driving while 'f***ing drunk' with a flat tire. *Mail Online*. https://www.dailymail.co.uk/news/article-3269963/Florida-woman-arrested-live-streams-video-driving-flat-tire-screaming-f-ing-drunk-motorists-heard-honking-her.html

Statista, (2018). 'Percentage of Internet Users Who Watch Online Video Content on Any Device as of January 2017, by Country'. https://www.statista

Straits Times, The. (2017, April 25). Thai man broadcasts baby daughter's murder live on Facebook. *The Straits Times*. https://www.straitstimes.com/asia/se-asia/thai-man-murders-child-kills-himself-on-facebook-live

Straits Times, The. (2019, May 20). North Korean women tell of paedophilia, slavery and gang rape on camera in Chinese cyber sex dens. *The Straits Times*. https://www.straitstimes.com/world/europe/north-korean-women-tell-of-paedophilia-slavery-and-gang-rape-on-camera-in-chinese-cyber

Strutin K. (2011). Social Media and the Vanishing Points of Ethical and Constitutional Boundaries. *Pace Law Review 31*, pp. 228–288.

Surette, R. (2012). 21st Century Crime and Justice, Social Media, and Maximizing Audience Participation. Pop Culture Universe: Icons, Idols, Ideas, ABC-CLIO. http://www.abc-clio.com

Surette, R. (2015). Performance Crime and Justice. *Current Issues in Criminal Justice, 27*(2), pp. 195–216.

Thailand shooting: How the massacre unfolded. (2020, February 09). *BBC News.* https://www.bbc.com/news/amp/world-asia-51430619

Whaley, S. (2017, June 5). Tulsa man accused of broadcasting child abuse on Facebook Live. *Fox 23 News*. https://www.fox23.com/news/tulsa-man-accused-of-broadcasting-child-abuse-on-facebook-live/530158051

Wheatstone, R. (2019, June 10). 'Predatory Paedo' Dad-of-two Managing Director, 43, is Jailed for Abusing Girl, 14, He Met on Snog.fm After She Outed Him at Family Dinner. *The Sun*. https://www.thesun.co.uk/news/9259906/managing-director-paedophile-outed-family-dinner/

Whitty, M. T., Doodson, J., Creese, S., & Hodges, D. (2017). A picture tells a thousand words: What Facebook and Twitter images convey about our personality. *Personality and Individual Differences*. Advance online publication.

Withers, T. (2019, May 3). Death Toll From New Zealand Mosque Terrorist Attacks Rises to 51. *Bloomberg*. https://www.bloomberg.com/news/articles/2019-05-02/new-zealand-death-toll

Woods, A. (2019, June 6). Monster gets 120 years in prison for livestreaming sexual abuse of his daughter. *New York Post*. https://nypost.com/2019/06/06/monster-gets-120-years-in-prison-for-live-streaming-sexual-abuse-of-his-daughter/

Worley, W. (2016, September 8). Father uses Facebook Live to confess shooting of ex-wife and son. *The Independent*. https://www.independent.co.uk/news/world/americas/facebook-live-earl-valentine-father-confess-shooting-wife-son-family-murder-a7231921.html

Yar, M. (2012). Crime, media and the will-to-representation: Reconsidering relationships in the new media age. *Crime Media Culture*, *0*, pp. 1–16.

Yuhas, A. (2017, January 20). Ohio mother who taped son to wall on Facebook Live faces charges. *The Guardian*. https://www.theguardian.com/technology/2017/jan/19/facebook-live-video-mom-tapes-son-wall-arrest

Yu, J., Chai, W., & Tambagan, B, J. (2020). Don't believe me, just watch: Live-streaming crime. *Home Team Journal*, *9*, pp. 44–56.

https://doi.org/10.1142/9789811232411_0002

# Chapter 2

# Hidden but Deadly: Stalkerware Usage in Intimate Partner Stalking

Stephanie Chan

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Stephanie_CHAN@mha.gov.sg*

## 2.1 Introduction

Yessenia Suarez was last seen at her mother's house in Deltona, Florida on the night of 22 October 2013. At the time of her disappearance, Suarez's two-year marriage to Luis Toledo was on the rocks. Earlier in the day, Toledo confronted Suarez at her workplace and slapped her, infuriated that she was having an affair with her co-worker. That very evening, they met up at Suarez's mother's house to talk. By the end of their discussion, both husband and wife agreed to end their marriage. They left the house separately, with Suarez driving herself and her children back home [The Charley Project, 2019].

The following morning, Suarez and her two children, Michael Otto and Thalia Otto, were reported missing. Toledo was later convicted of second-degree murder in Suarez's death, and of first-degree murder in each of the children's deaths [Fernandez, 2018a]. To date, their bodies were never found.

Court proceedings revealed that the couple's marriage was fraught with difficulties [Ganney, 2017]. The couple argued constantly about family responsibilities. Toledo was also allegedly physically violent to his

wife, and a family relative had said that the couple was "always fighting." Suspicious of his wife's infidelity, Toledo secretly installed spyware on her phone to gain remote access to her phone messages. In his confession, he said that he had discovered the affair through the spyware and killed her with "a strike to her throat."

### 2.1.1 *The call to look at* Technology *misuse in intimate partner stalking*

In the case of the Deltona triple murder, Yessenia Suarez had a partner who was physically abusive towards her and utilised spyware to monitor her mobile phone communication with others [Fernandez, 2018b]. In other words, Luis Toledo had abused technology and carried out violence and stalking behaviours against his spouse.

Intimate Partner Stalking (IPS) is closely associated with abusive relationship behaviour. Also termed as relational stalking, IPS occurs in various forms, including surveillance, confrontations, repeated communication attempts, and threats. Non-profit social organisation Peace Over Violence likens it to a form of "coercive control where one person attempts to exert power over another" [PeaceOverViolence, n.d.]. IPS is notedly present in stages where relationship deterioration or dissolution has occurred [Mechanic *et al*., 2002]. Much of the literature examines IPS in the post-breakup stage of a relationship [Ferreira & Matos, 2013]. In a survey of 305 female British undergraduates who had been in heterosexual romantic relationships [Roberts, 2005], a considerable number of participants (34%) self-reported that they had been a victim of stalking following the termination of a romantic relationship. In Singapore, a recent survey[a] of 226 women found that half of them (54.9%) have experienced a stalking incident [Sheridan *et al*., 2018].

In addition to the pervasiveness of IPS in abusive relationships, there is also a growing concern about the misuse of technology—ranging from smartphone apps to Internet-connected home devices—by intimate partner abusers [Bowles, 2018]. In this chapter, we will examine the use of smartphone spyware in the context of IPS (i.e., stalkerware usage), with a focus on stalkerware as the ultimate stalking tool. The prevalence of

[a]Published studies of IPS prevalence in Singapore are few, with researchers (Sheridan *et al*., 2018) suggesting a need for greater public awareness.

stalkerware in IPS, the perpetrator's mindset in using stalkerware, and recommendations to alleviate the harm caused to victims will also be discussed.

## 2.2 The Rise of Stalkerware as the Ultimate Stalking Tool

Driven by increasing smartphone ownership and advancing communication technologies, IPS is gaining an impact in this digital age. Smartphone penetration rates have been on the rise, from a 33.5% global penetration rate in 2016 to 41.5% penetration rate in early 2019 [Holst, 2019]. It is projected that the world will see 3.2 billion smartphone users in early 2020 (i.e., an estimated 44.9% of the world's population). For developed countries, the penetration rates are staggering. According to the Deloitte 2019 global mobile consumer survey, nine out of 10 survey respondents in developed countries own a smartphone [Lee *et al*., 2019]. As more people use or own a smartphone, and as the carrying of smartphones becomes increasingly normalised, they likewise become accustomed to the device's embedded functions in their personal, social, and professional lives.

Smartphone technology is also rapidly advancing. The variety of smartphone applications (apps) and accessories available in the market enables smartphones to perform numerous functions for the modern individual. For instance, the three most commonly used types of apps are social media apps, gaming apps, and communication apps [Panko, 2018].[b] Such forms of integrated technology solidify the use of smartphone apps. At the same time, consumers become accustomed to the seamless incorporation of smartphones into their daily routines [Lee *et al*., 2019].

On the flip side, this raises the concern that a diverse and excessive amount of personal data can be obtained from an individual via different smartphone apps [Isaac, 2011], including:

- Communication content and usage history
- Geographical locations at given points in time

[b]Apps can also be installed in tablets and similar devices, but for conciseness, this paper assumes that stalkerware usage in smartphones is similar to their usage in other similar smart devices.

- Fitness activities
- Online transaction records
- Usage patterns of "smart" home appliances
- Mobile banking access
- Social networking access

### 2.2.1 *Defining spyware, stalkerware, and dual-use technology*

Unfortunately, a conducive digital environment also encourages a rise in the development and/or abuse of smartphone-supported apps that facilitate IPS and perpetuate and accentuate its effects. At this point, it is important to note the various relevant terms that are used when discussing stalkerware, namely, spyware, stalkerware, and dual-use technology. Several clear definitions have been proposed by The Citizen Lab, a multidisciplinary laboratory based at the University of Toronto [Parsons *et al*., 2019].

*Spyware* is defined as "software that enables a remote user to covertly obtain data about another individual's activities on an electronic device by surreptitiously transmitting data from the targeted device to another computer system" [Parsons *et al*., 2019, p. 10]. Covert data transmission is deemed essential in many areas, including that of the activities of a nation, an organisation, or a person.

*Stalkerware* refers to a category of spyware that is developed specifically to assist in intimate partner stalking. Within the context of IPS, it is defined as "applications that exist or may be installed on a mobile device, to let the operator of the application remotely monitor the activities of the device's user, or individuals routinely in the proximity of the user, … (which) includes intimate partner spyware applications and (general purpose) spyware applications which are repurposed to facilitate intimate partner violence, abuse, and harassment" [Parsons *et al*., 2019, p. 10].

*Dual-use technology* is defined broadly as "technology that may be intended for, or may be used for, legitimate or benevolent ends, but which may be equally capable of being repurposed for illegal, harmful, or unethical practices" [Parsons *et al*., 2019, p. 10]. In the context of IPS, stalkerware is easily accessible when perpetrators repurpose apps meant for parental monitoring or employee monitoring. For example, the parental monitoring app "PhoneSheriff" has capabilities for covert access to the target's texts, photos, and GPS location [Curtis, 2019]. Even trusted

phone location service apps, such as 'Find My iPhone' can be misused to facilitate IPS [Dzikiy, 2018].

Putting all three terms together, stalkerware are downloadable smartphone apps, readily available from Internet app stores or privately-owned websites, with customised spyware capabilities that enable the perpetrator to covertly gather data from a target's phone. These overlapping definitions between stalkerware and spyware showcase the versatility of app usage and hint at the murky opportunistic nature of app users [Parsons *et al*., 2019].

### 2.2.2 *Stalkerware's invasive capabilities*

Stalkerware allows for discreet round-the-clock tracking and monitoring. According to The Citizen Lab, common stalkerware capabilities that are advertised by app developers and/or app distributors online include the following [Parsons *et al*., 2019]:

(1) Record, access, or monitor GPS, keystrokes, calendar, contacts, email, web traffic, stored media, social media, phone logs, chat apps, SMSes, and phone calls
(2) Access backup data
(3) Block phone calls
(4) Authorise update of hidden stalkerware apps
(5) Activation of phone camera and/or phone microphone.

Stalkerware capabilities clearly entice IPS perpetrators to select them as the modern-day surveillance tool. This is especially true since stalkerware is more intrusive and "data rich" than the older forms of surveillance technology installed in desktop computers, vehicles, and houses. In fact, there is growing concern regarding the use of stalkerware and/or spyware in cases of deteriorating romantic relationships [Hodgson, 2019; Tidy, 2019]. The next section will discuss more about the rise of stalkerware.

## 2.3 Growing Global Trends of Stalkerware in IPS

Research on the phenomenon of stalkerware usage in IPS is still relatively new, given that the first iPhone was released by Apple in mid-2007, and

the Apple App store was opened in mid-2008. Similarly, the first Android smartphone and its accompanying Android Market (i.e., the present-day Google Play) were made available only in late 2008. A mere decade of smartphone technological advancement meant that current literature and research surveys on the prevalence and nature of stalkerware are still exploratory in nature.

### 2.3.1 *Major studies conducted on stalkerware prevalence: The 2013 SmartSafe project and the 2015 ReCharge survey*

One of the key studies conducted on stalkerware prevalence is the SmartSafe project, funded by the Victoria Legal Aid and carried out by the Domestic Violence Resource Centre Victoria (DVRCV),[c] as part of a large-scale examination of technology-facilitated domestic stalking against women [Woodlock, 2013]. The first survey in the project was conducted with 152 workers in the domestic violence sector from Victoria, Australia. Survey findings showed that most workers (97%) noticed the use of mobile technologies to stalk women, and that there were a variety of technologies used in stalking: smartphones (82%), social media (82%), email (52%), and GPS tracking (29%). Given that Australia's smartphone penetration rate continued to rise steadily from 43.7% in 2012 to 68.3% in 2017 [Hughes, 2020], the use of stalkerware in IPS is likely to increase.

The second survey in the SmartSafe project was conducted with 46 victims with an average age of 35. Survey findings showed that some victims had partners who set up their phone for them (22%) and had partners who had downloaded apps onto their phone (20%). The victims also reportedly shared their phone passwords with others (22%), with only half knowing how to amend some of their security settings (49%). More worryingly, out of 44 respondents, half had a partner who utilised mobile technology to check where they were (56%), and some had a partner who tracked them with GPS apps (17%). There were other forms of technology-based abuse, such as partners who demanded their electronic

[c]A state-wide resource centre that provides training, publications, research and other resources to those experiencing family violence.

passwords (17%), partners who had impersonated them on email, text messages, and/or social media (14%), and partners who had purchased a phone for the victim for the purpose of keeping track of them (8%). In summary, the SmartSafe project utilised the perspectives of two groups—workers in the domestic violence sector and victims of domestic stalking—to show that smartphones and mobile apps are quickly becoming the preferred stalking tool of IPS perpetrators.

Building on the 2013 SmartSafe project, the ReCharge Survey [Woodlock, 2015] further examined this ongoing trend of digital-facilitated domestic stalking, in particular, the abuse of smartphone apps. Funded by the Australian Communications Consumer Action Network (ACCAN), this survey was a project collaboration between the DVRC, the Women's Legal Service NSW, and the Women's Services Network (WESNET). The survey's reach extended beyond the state of Victoria to the other states, making it a national study for Australia. A total of 546 workers in the domestic violence sector[d] had participated. Results show that almost all practitioners (98%) have had clients who experienced technology-facilitated stalking. Common types of phone technologies abused were: text messaging (82%), Facebook (82%), and GPS tracking via apps (29%). Of particular concern is the fact that 34% of practitioners reported it was common for them to see smartphone app abuse for GPS tracking in their cases (i.e., "all the time"). The survey findings also showed that information obtained through smartphone technology was used by perpetrators to engage in verbal abuse and verbal threats towards their victims, and to heighten perpetrators' monitoring efforts of the victim's behaviour. Regarding verbal threats, some practitioners had cases in which the perpetrator threatened distribution of private photos and videos of victims (i.e., "often" (35%) and "all the time" (14%)). Ultimately, the survey highlighted the significant rise of this phenomenon and urged for more extensive research to explore solutions for prevention, intervention, and victim-support.

---

[d]The project's objectives included examining the prevalence rate, the types of digital abuse, as well as practitioner perspectives on the effectiveness of both legal and police responses. The practitioners were mainly working in domestic violence organisations (53%), while the rest worked in legal organisations, sexual assault cases, housing cases, and/or health organisations.

### 2.3.2 *Usage trends in UK and USA*

Stalkerware usage in IPS is also an issue in the UK and the US. The Women's Aid Federation of England (i.e., Women's Aid), a charity against domestic violence, considers the use of stalkerware and spyware in intimate relationships as "digitally assisted stalking" and emphasises the importance of digital safety for victims [Perry, 2012]. A recent study in New York City conducted focus groups with 39 victims and interviews with 50 professionals in the domestic violence sector [Freed *et al*., 2018]. This qualitative study, in collaboration with the New York City Mayor's Office to Combat Domestic Violence (OCDV) and five family justice centres, sought to understand the prevalence rate and attack strategies of technology-facilitated abuse in intimate relationships. The findings from the interviews with practitioners showed that it was common for perpetrators to install tracking apps onto their victims' phones (n = 47). This includes the misuse of legitimate tracking apps such as 'Find My Phone', an anti-theft app, and 'Find My Friends', an emergency response app. Findings from the focus group also showed that while it was common for victims to suspect the presence of spyware or dual-use apps (n = 15), few were able to confirm the presence of such spyware (n = 3). Many of the victims' phones were compromised by a combination of their partners employing persuasion tactics or physical threats to obtain the phone password (n = 28), going through the victim's unlocked phone when the victim was unaware (n = 16), or inferring the phone passwords and the answers to security questions (n = 26). Above all, the results of this qualitative study supported the findings of both the SmartSafe surveys, which showed that the stalkerware phenomenon is a worrying trend that underlies the turbulent dynamics of abusive intimate relationships.

### 2.3.3 *Distributions and download trends in global online stores*

Globally, the demand for stalkerware persists. In Nanjing, China, police investigators traced a stalkerware developer who allegedly sold the app to 40 other domestic agents [Abacus, 2019]. Kaspersky, a Russian cybersecurity provider, reported findings in 2019 that there were more than 37,000 unique users whose anti-spyware programme detected stalkerware in their smartphones at least once, a significant 35% increase from 27,000 users in 2018 [Kaspersky, 2019]. In their report, Russia

accounted for 25.6% of these users, followed by India (10.6%), Brazil (10.4%), US (7.1%), Germany (3.6%), Italy (2.7%), Mexico (2.1%), UK (2.0%), France (1.7%), and Iran (1.7%). Avast, another cybersecurity provider with its headquarters in the Czech Republic, reported findings in 2019 that eight stalkerware apps[e] were detected on the Google Play store, of which the combined number of installations on smartphones was 140,000 times [Elder, 2019]. Two out of these eight apps, 'Spy Tracker' and 'SMS Tracker', each had at least 50,000 installations. All apps were reported by Avast to Google Play and subsequently removed.

Clearly, there has been a growing awareness among researchers, domestic violence charities, law enforcement agencies, and cybersecurity providers of the thriving commercial demand for stalkerware. Stalkerware developers are likewise catering to the growing consumer demand for powerful tracking apps that allow the user to gain unrestricted control over a target's smartphone with minimal effort and cost. Nevertheless, there is still a need to understand the psyche of perpetrators who utilise stalkerware to perpetuate IPS.

## 2.4 Mindset of the Stalkerware Perpetrator

How can we make sense of the mindset of an IPS perpetrator who covertly installs stalkerware on a victim's smartphone?

From the perspective of victims and workers in the domestic violence sector, perpetrators employ smartphone technology to cause fear, to create a sense of omnipresence, and to punish and humiliate [Woodlock, 2013]. Some practitioners have observed that perpetrators use stalkerware to carry out abuse, to threaten, and to impersonate [Woodlock, 2015]. From the perspective of stalkers who targeted their current or former intimate partners, their stalking behaviours were either in response to a relationship rejection or a desire for relationship reconciliation, with self-reported emotions of anger, frustration, jealousy, sadness, and a sense of loss [Dardis & Gidycz, 2019; Mullen *et al*., 1999]. Interestingly, a study of 187 women who have been stalked by former partners [Brewster, 2003] found that relationship reconciliation was the primary purpose for initial stalking

[e]The eight apps were sold under the following names: 'Track Employees Check Work Phone Online Spy Free', 'Spy Kids Tracker', 'Phone Cell Tracker', 'Mobile Tracking', 'Spy Tracker', 'SMS Tracker', 'Employee Work Spy', and 'Family Employee Monitor'.

behaviours (75%). However, if reconciliatory efforts were unsuccessful, the victims felt that their stalkers' intents broadened out into desires to enact revenge (45%) and to regain control (27%). Such is the changing dynamic of stalking and surveillance in intimate relationships.

A greater understanding of the underlying psychological mechanisms that drive the varied desires of stalkerware users would be helpful in combating stalkerware usage. Although stalkerware is a relatively recent phenomenon with limited studies, there are clear parallels with relational stalking and cyberstalking. Therefore, our understanding of stalkerware usage can be drawn from the existing literature to determine the factors that drive such behaviours. They are namely: unhealthy attachment styles, coercive control, and moral disengagement.

### 2.4.1 *Unhealthy attachment styles*

While stalkerware provides the app user with rich information about a partner, much of this information is frequently gained without the target's knowledge or consent. The need for secrecy or the use of force to facilitate stalkerware downloads onto a target's phone posits that stalkerware usage potentially occurs in an intimate relationship that has an unhealthy interaction dynamic.

Attachment theory [as cited in Firestone, 2013] posits that the early bond between a child and the caregiver results in an attachment style that persists into adulthood relationships, namely secure attachment, anxious attachment (e.g., preoccupied by rejection and abandonment), avoidant attachment (e.g., emotionally distant and dismissive), and disorganised attachment (e.g., unresolved losses from prior trauma). In particular, the development of either an anxious attachment style or an avoidant attachment style is linked to the experience of more aggressive behaviours and interpersonal problems in intimate relationships [Bookwala & Zdaniuk, 1998].

The effects of insecure attachment styles on stalkers are most often studied in the post-relationship stage. In particular, the anxious attachment style is linked to greater desire for relationship reconciliation [Johnson & Thompson, 2016]. The authors posited that high attachment anxiety causes individuals to have a greater fear of rejection and difficulty in accepting relationship dissolution. This is in line with previous studies showing that high anxiety in attachment styles is linked to the use of a greater number of unwanted pursuit behaviours (UPBs) [Tassy &

Winstead, 2014], an increased preoccupation with the ex-partner via rumination [De Smet *et al*., 2015], and greater emotional distress over real or perceived unsuccessful relationships [MacKenzie *et al*., 2008]. Dardis and Gidycz's 2019 study on individuals who engaged in direct and cyber UPBs found that difficulty in regulating emotions relating to the breakup was linked to having a high anxiety construct in their attachment styles.

Thus, an anxious attachment style (i.e., one of the insecure attachment styles) is a possible mechanism that underlies the perpetrator's persistent use of stalkerware. The effectiveness of stalkerware in tracking and extracting conversation information from phone calls and chat apps renders it extremely useful for individuals who are suspicious of possible relationship deterioration and dissolution.

### 2.4.2 *The need for coercive control*

Next, stalkerware also enables the app user to gain knowledge about the victim's whereabouts and their interactions with others. Often, much of this information is used by the app user to exert cyber dominance and control over the victim.

The literature on IPS agrees that abusive relationship behaviours are driven by the abuser's need for power and control [Brewster, 2003]. The Coercive Control Approach posits that coercive control is exhibited by an unceasing use of intimidation, degradation, isolation, and control, which may or may not always include physical violence [Stark, 2007]. The earliest interpretation by Johnson [2008, as cited in Meier, 2016] explored the typologies of domestic violence from a feminist perspective, outlining that coercive control (termed as 'intimate terrorism') occurs when the more powerful abuser (mainly male) attempts to dominate the victim (mainly female) by exerting control through the use of control tactics. The more recent interpretation allows that coercive control tactics need not always involve physical violence, and may involve monitoring and regulation of daily activities; non-compliance results in harmful consequences for the victim [Crossman & Hardesty, 2017; Stark, 2012].

Coercive control is expressed in various forms—isolation, micro-regulation, and a sense of imprisonment—across the physical, emotional, social, and financial areas of a victim's life [Brewster, 2003]. Isolation occurs by controlling access and limiting the victim's contact with their family and their broader social network. Micro-regulation occurs by

requiring compliance with a closely monitored schedule of daily activities. A sense of imprisonment occurs when there is a persistent combination of isolation, micro-regulation, manipulation, and fear that the victim feels unable to break out of [Crossman & Hardesty, 2017]. Therefore, coercive control is a possible mechanism that underlies the perpetrator's initial interest in and subsequent persistent use of stalkerware. In particular, the tracking and monitoring capabilities allows for micro-regulation and isolation tactics to be carried out extensively, thus subjecting a victim under the perpetrator's control.

### 2.4.3 *Use of moral disengagement to perpetuate the act*

Unsurprisingly, stalkerware with the capabilities to mask their presence and/or real purposes in a victims' smartphone are highly sought after in the online market. Considering how perpetrators gain access to their target's smartphone via covert or deceptive means, it is possible that the act of app monitoring requires cognitive justification efforts.

Based on the advertising content of various stalkerware websites [Chatterjee *et al*., 2018; Parsons *et al*., 2019], it is likely that stalkerware users engage in moral disengagement when carrying out their stalking behaviours. Bandura [1999] defines moral disengagement as "the cognitive restructuring of inhumane conduct into a benign or worthy one" [p. 1]. In other words, moral disengagement is a process of convincing oneself that ethical standards do not personally apply in a particular context. Moral Disengagement theory provides for seven practices in which moral disengagement can occur [Bandura, 2002], many of which can be applied to explain stalkerware purchases:

| **The Seven Moral Disengagement Variants** | **Example in Stalkerware for IPS context** |
|---|---|
| *Moral Justification*: A means of restructuring the behaviour to make it morally or socially acceptable. | App users might see themselves as trying to prevent relationship dissolution. For example, stalkerware FlexiSPY states it can "protect your relationships." |
| *Euphemistic Labelling*: A means of using language to make the behaviour appear respectable. | App users might be swayed by the marketing tactics that promote apps using seemingly respectable motives. For example, an infamous stalkerware calls itself "TheTruthSpy." |

| The Seven Moral Disengagement Variants | Example in Stalkerware for IPS context |
|---|---|
| *Advantageous Comparison*: A means of employing the contrast principle to downplay the harm. | App users might be susceptible to marketing tactics that highlight the benefits of surveillance apps. For example, stalkerware websites include customer testimonials of how the app assisted couples in solving relationship problems. |
| *Displacement of Responsibility*: A means of shifting the blame to another source, usually to a higher authority. | No known application by stalkerware perpetrators. |
| *Diffusion of Responsibility*: A means of reducing personal contribution. | App developer companies have policies that place liability on the app user. For example, the stalkerware Hoverwatch's policy statement includes the sentence "by installing the software or using the service you certify that you act in accordance to the law and you take full responsibility for the use of the project."<br>There is no known application by stalkerware perpetrators. |
| *Disregard or Distortion of Consequences*: A means of ignoring or reducing the harmful impact. | App users might find it easier to control and intimidate their victims since suffering and psychological trauma is muted or not immediately visible in the online world. For example, the absence of physical harm to the victim might give the misconception that there is no harm inflicted. |
| *Dehumanisation (of Victim)*: A means of taking away the human qualities from people. | No known application by stalkerware perpetrators. |

In summary, while research on stalkerware is still in its infancy stages, insights can be from the literature on relational stalking, cyber stalking, and deviance due to the similarities in psychological elements. As shown above, the mindset of the stalkerware perpetrator comprises various motivations driven by three underlying mechanisms—unhealthy attachment styles, need for coercive control, and the use of moral disengagement to perpetuate stalkerware usage in IPS. By observing the

underlying psychological mechanisms of stalkerware perpetrators, we can see how app usage is often just an amplified expression of the user's deep-seated relationship attitudes and perceptions.

## 2.5 Alleviating the Harm: Barriers and Recommendations

Much needs to be done in terms of stalkerware-facilitated IPS prevention, deterrence, awareness, and the alleviation of harm to affected individuals. Significantly, many victims self-reported in the SmartSafe survey that their mental health and wellbeing were affected across a wide range of areas including sleeping habits, daily routines, work/academic performance, parenting, and social networks [Woodlock, 2013]. Practitioners in the ReCharge survey [Woodlock, 2015] also shared that their clients experienced hyper-vigilance, a constant sense of fear, social isolation, and a perception that the perpetrator was omnipresent. In addition, many of these cases eventually escalated into incidents of doxing, revenge pornography, non-consensual sexting, blackmail, public humiliation, harassment, and physical threats [Woodlock, 2015].

The numerous underlying social factors contributing to the phenomenon of stalkerware means that there, too, are many existing barriers to effective interventions. Therefore, a multi-pronged approach should be adopted in order to overcome these barriers[f].

### 2.5.1 *Overcome victim-blaming*

According to the "just world" hypothesis, victims of crime are held somewhat accountable for their injuries, as a way of explaining crime occurrence [Mancini & Pickett, 2017]. The same is seen in the literature on stalking. A study in an Australian university showed that victims of intimate stalking vignettes were viewed as more responsible for the stalking situations as compared to victims of stranger stalking vignettes [Scott *et al*., 2014]. Likewise, a mock-jury study found that fewer guilty verdicts were handed down when the stalking was conducted

[f]These approaches must still be calibrated according to the IPS trend in the local or regional context and to the extent of reach and resources by the different organisations and agencies.

electronically as compared to a face-to-face situation [Magyarics *et al*., 2015]. The mock jurors had interpreted that electronic stalking did not highlight a stalker's intention to create fear in the victim, nor were they fully convinced that it negatively impacted the victims. Unsurprisingly then, victim-blaming is equally rife in the context of stalkerware. In fact, it is common for stalkerware advertisements to depict the intimate partners and spouses as legitimate targets of tracking and monitoring apps due to their purported misdeeds (e.g., "catch a cheating spouse") [Parsons *et al*., 2019]. Victim-blaming also has negative implications as it further isolates the victim as well as reduces victim reporting rates.

#### 2.5.1.1 *Recommendations to overcome victim-blaming*

| | |
|---|---|
| **Support for Victims** | #1: Professionals and counsellors to work with their clients, if need be, to help victims to respond appropriately to criticisms directed to them by their loved ones or close others.<br>#2: Investigative officers, when conducting investigative interviews, to look out for signs of victim-blaming, and to direct victims to appropriate counselling resources when necessary.<br>#3: Forum administrators and moderators to look out for instances of victim-blaming, and to respond swiftly and appropriately to all forum parties involved. |
| **Public Awareness and Education** | #4: Carry out targeted ad campaigns to raise public awareness on prevalence rates, to evoke empathy towards victims, and to educate on myths surrounding IPS. Such campaigns can be run by relevant local organisations, law enforcement agencies, or even by technology companies on an international level. |

### 2.5.2 *Help victims to tackle proxy stalking*

While victims might be aware of potential stalkerware in their own smartphones, perpetrators can circumvent this by installing stalkerware in the smartphones and mobile devices of the victim's family members to facilitate proxy stalking [Freed *et al*., 2018]. In a survey of 346 victim service providers, programmes, or agencies by the National Network to End Domestic Violence (NNEDV), more than half (60%) reported perpetrators who spied on the children of their targeted partners [NNEDV, 2014]. The survey also revealed that children's smartphones were the device-of-choice (89%), of which the perpetrators either installed the

stalkerware onto the child's smartphone or gave a compromised smartphone as a gift to the child.

#### 2.5.2.1 *Recommendations to help victims to tackle proxy stalking*

| | |
|---|---|
| **Support for Victims and their Families** | #1: Anti-stalking organisations to raise awareness of proxy stalking through their websites and their campaign materials.<br>#2: Professionals and counsellors to raise awareness of proxy stalking to their clients, and if need be, to guide them in taking the necessary steps to protect their devices and those owned by their children.<br>#3: Shelters for victims of IPS or domestic violence to advise victims accordingly on how to handle instances of real or suspected proxy stalking. |
| **Technology Protection in the App Markets** | #4: Anti-spyware technology companies to update the capabilities of their anti-spyware apps to enhance mobile device protection.<br>#5: If possible, for a coordinated effort between domestic shelters and app development companies to provide subsidised anti-spyware for victims of proxy stalking cases. |
| **Legal Recourse for Victims** | #6: Legal amendments to existing harassment or stalking laws, depending on the country, to take into account the negative impact on victims of proxy stalking means. At present, the laws and legal coverage for harassment, stalking, and domestic disputes differ across countries. |

### 2.5.3 *Improve detection by anti-spyware*

Cohen's Routine Activity Theory [as cited in Miro, 2014] states that crime occurs when a motivated offender comes into contact with a desirable and suitable target in the absence of capable guardians. Anti-spyware are guardians against secretly installed stalkerware and/or tracking apps repurposed into stalkerware. Unfortunately, anti-spyware programmes are limited in their ability to detect stalkerware. Chatterjee and colleagues [2018] tested 40 anti-spyware apps on their ability to detect 276 in-store spyware apps (either overt stalkerware or dual-use apps that can potentially be repurposed as stalkerware) and 20 off-store spyware apps. They found that most of the anti-spyware apps (n = 37) failed to detect more than 3% of the combined in-store and off-store spyware. The more effective anti-spyware apps were also found to be better at detecting

off-store spyware as compared to in-store spyware. It was posited that anti-spyware apps might lack rigorous standards in identifying stalkerware, with anti-spyware developers being reluctant to consider in-store dual-use apps as having the potential for being repurposed as stalkerware.

#### 2.5.3.1 *Recommendations to improve detection by anti-spyware*

| | |
|---|---|
| **Technology Protection in the App Markets** | #1: Anti-spyware technology companies to constantly review and assess the capabilities of their anti-spyware programmes for detecting both in-store and off-store stalkerware. These findings should be published publicly.<br>#2: A joint effort between the key players in the anti-spyware industry to share technologies to lower app development costs, such that affordable anti-spyware programmes can be developed for victims.<br>#3: Major app stores (e.g., Google LLC, Apple Inc.) to constantly review the apps in their stores, to discover and remove stalkerware, as well as to detect dual-use apps that put up IPS-related advertisements (e.g., an advertisement that promotes the use of an app to "catch a cheating spouse"). |
| **Target Stalkerware Distributors** | #4: If possible, an international effort between the anti-spyware industry, the major app stores (e.g., Google LLC, Apple Inc.) and law enforcement agencies to consolidate data on existing stalkerware apps, download data, and purchase data. The data can be used to (i) triangulate the detection and apprehension of regional stalkerware distributors, (ii) identify characteristics of stalkerware consumers, and (iii) shut down the websites that allow off-store stalkerware downloads. |

### 2.5.4 *In sum: A concerted effort by all stakeholders*

Ultimately, addressing the problem of stalkerware usage requires a concerted effort from all stakeholders, particularly in reporting and removing stalkerware apps, so that such smartphone technologies are less accessible to potential perpetrators. For stakeholders in the cyber security sphere, a unified approach towards improving surveillance detection tools would be beneficial in supporting victims. At the same time, victim support agencies can update their advice and awareness materials to raise

awareness of stalkerware-facilitated IPS, and to provide better support for victims of (actual or suspected) stalkerware perpetration.

## 2.6 Conclusion

In summary, stalkerware-facilitated IPS is a new but rising phenomenon. The use of stalkerware and its related smartphone technologies (i.e., spyware and repurposed dual-use apps) in IPS has led to greater and more destructive mental harm on victims. Although research on the psychology of perpetrators has shed light on the pervasive use of stalkerware, literature on stalkerware is still in its exploratory stages. Perhaps with the combination of continued research and concerted stakeholder efforts, there would be marked improvements in prevention, deterrence, and the alleviation of harm to victims in the near future.

## 2.7 Acknowledgement

The views expressed in this chapter are the author's only and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 2.8 References

Abacus (2019, November 6). Police in China uncover 'stalkerware' app that lets users spy on partners. *South China Morning Post*. https://www.scmp.com/tech/big-tech/article/3036612/police-china-uncover-stalkerware-app-lets-users-spy-partners?li_source=LI&li_medium=homepage_hk_edition_top_picks_for_you

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*, pp. 193–209. https://doi.org/10.1207/s15327957pspr0303_03

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education, 31*, pp. 101–119. https://doi.org/10.1080/0305724022014322

Bookwala, J., & Zdaniuk, B. (1998). Adult attachment styles and aggressive behaviour within dating relationships. *Journal of Social and Personal Relationships, 15*, pp. 175–190. https://doi.org/10.1177/0265407598152003

Bowles, N. (2018). Thermostats, locks and lights: Digital tools of domestic abuse. *The New York Times*. https://www.nytimes.com/2018/06/23/technology/smart-home-devices-domestic-abuse.html

Brewster, M. (2003). Power and control dynamics in prestalking and stalking situations. *Journal of Family Violence, 18*, pp. 207–217. https://doi.org/10.1023/A:1024064214054

Chatterjee, R., Doerfler, P., Orgad, H., Havron, S., Palmer, J., Freed, D., Levy, K., Dell, N., McCoy, D., & Ristenpart, T. (2018). In conference. IEEE Symposium on Security and Privacy. San Francisco, USA.

Crossman, K., & Hardesty, J. (2017). Placing coercive control at the center: What are the processes of coercive control and what makes control coercive? *Psychology of Violence, 8*, pp. 196–206. https://doi.org/10.1037/vio0000094

Curtis, C. (2019). Domestic abusers are using easily accessible apps to stalk and control their victims. *The Next Web*. https://thenextweb.com/codeword/2019/07/16/domestic-abusers-are-using-easily-accessible-apps-to-stalk-and-control-their-victims/

Dardis, C., & Gidycz, C. (2019). Reconciliation or retaliation? An integrative model of postrelationship in-person and cyber unwanted pursuit perpetration among undergraduate men and women. *Psychology of Violence, 9*, pp. 328–339. https://doi.org/10.1037/vio0000102

De Smet, O., Uzieblo, K., Loeys, T., Buysse, A., & Onraedt, T. (2015). Unwanted pursuit behaviour after breakup: Occurrence, risk factors, and gender differences. *Journal of Family Violence, 30*, pp. 753–767. https://doi.org/10.1007/s10896-015-9687-9

Dzikiy, P. (2018). Tracking or stalking? The dark side of tracking apps. *Security Baron*. https://securitybaron.com/blog/tracking-or-stalking-the-dark-side-of-tracking-apps/

Elder, J. (2019). Google pulls stalker apps identified by Avast. *Avast Blog*. https://blog.avast.com/avast-identifies-stalker-apps

Fernandez, Frank. (2018a, November 7). Killer keeps secret where he hid wife, children. *News Herald*. https://www.newsherald.com/news/20181107/killer-keeps-secret-where-he-hid-wife-children

Fernandez, Frank. (2018b, October 12). Ex-lover testifies about final conversation before Deltona mom disappeared. *The Daytona Beach News-Journal*. https://www.news-journalonline.com/news/20171012/ex-lover-testifies-about-final-conversation-before-deltona-mom-disappeared

Ferreira, C., & Matos, M. (2013). Post-relationship stalking: The experience of victims with and without history of partner abuse. *Journal of Family Violence*, *28*, pp. 393–402. https://doi.org/10.1007/s10896-013-9501-5

Firestone, L. (2013). How your attachment style impacts your relationship. *Psychology Today.* https://www.psychologytoday.com/sg/blog/compassion-matters/201307/how-your-attachment-style-impacts-your-relationship

Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T., & Dell, N. (2018). A stalker's paradise: How intimate partner abuses exploit technology. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Paper No. 667, pp. 1–13. https://doi.org/10.1145/3173574.3174241

Ganney, Michelle (2017, October 20). Accused triple murderer confesses killing wife. *DailyMail Online*. https://www.dailymail.co.uk/news/article-5001236/Accused-triple-murderer-confesses-killing-wife.html

Hodgson, C. (2019). Inside the secretative world of stalking apps. *Financial Times*. https://www.ft.com/content/263133ac-a28b-11e9-974c-ad1c6ab5efd1

Holst, A. (2019). Global smartphone penetration rate as share of population from 2016 to 2020. *Statista*. https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/

Hughes, C. (2020). Smartphone penetration rate as share of the population in Australia from 2012 to 2022 (projection). *Statista*. https://www.statista.com/statistics/321477/smartphone-user-penetration-in-australia/

Isaac, M. (2011). Survey finds smartphone apps store too much personal data. *Wired*. https://www.wired.com/2011/08/smartphone-local-data-storage/

Johnson, E., & Thompson, C. (2016). Factors associated with stalking persistence. *Psychology, Crime & Law, 22*, pp. 870–902. https://doi.org/10.1080/1068316X.2016.1197225

Kaspersky (2019). The state of stalkerware in 2019. *Securelist*. https://securelist.com/the-state-of-stalkerware-in-2019/93634/

Lee, P., Casey, M., Wigginton, C., & Calugar-Pop, C. (2019). Deloitte's 2019 global mobile consumer survey: Tracking consumers' digital behaviour around the world. *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/industry/telecommunications/global-mobile-consumer-survey.html

MacKenzie, R., Mullen, P., Ogloff, J., McEwan, T., & James, D. (2008). Parental bonding and adult attachment styles in different types of stalker. *Journal of Forensic Science, 53*, pp. 1,443–1,449. https://doi.org/10.1111/j.1556-4029.2008.00869.x

Magyarics, C., Lynch, K., Golding, J., & Lippert, A. (2015). The impact of frequency of behavior and type of contact on judgments involving a criminal stalking case. *Law and Human Behavior, 39*, pp. 602–613. https://doi.org/10.1037/lhb0000151

Mancini, C., & Pickett, J. (2017). Reaping what they sow? Victim-offender overlap perceptions and victim blaming attitudes. *Victims & Offenders, 12*, pp. 434–466. https://doi.org/10.1080/15564886.2015.1093051

Mechanic, M. B., Weaver, T. L., & Resick, P. A. (2002). Intimate partner violence and stalking: Exploration of patterns and correlates in a sample of acutely battered women. In Davies, K. E., Frieze, I. H., & Maiuro, R. D. (Eds), *Stalking: Perspectives on victims and perpetrators,* (New York: Springer Publishing Company). pp. 62–88.

Meier, J. (2016). Differentiating domestic violence types: Profound paradigm shift or old wine in new bottles? *Domestic Violence, Abuse, and Child Custody: Legal Strategies and Policy Issues, 2*, pp. 7–35.

Miro, F. (2014). Routine activity theory. *The Encyclopaedia of Theoretical Criminology*, First Edition. Ed. J. Miller. Blackwell Publishing.

Mullen, P., Pathe, M., Purcell, R., & Stuart, G. (1999). Study of stalkers. *American Journal of Psychiatry, 156*, pp. 1,244–1,249. https://doi.org/10.1176/ajp.156.8.1244

NNEDV (2014). A glimpse from the field: How abusers are misusing technology. *The Safety Net Project*. https://www.techsafety.org/blog/2015/2/17/a-glimpse-from-the-field-how-abusers-are-misusing-technology

Panko, R. (2018). Mobile app usage statistics 2018. *The Manifest*. https://themanifest.com/app-development/mobile-app-usage-statistics-2018

Parsons, C., Molnar, A., Dalek, J., Knockel, J., Kenyon, M., Haselton, B., Khoo, C., & Deibert, R. (2019). The predator in your pocket: A multidisciplinary assessment of the stalkerware application industry. *The Citizen Lab*. https://citizenlab.ca/2019/06/the-predator-in-your-pocket-a-multidisciplinary-assessment-of-the-stalkerware-application-industry/

PeaceOverViolence (n.d.). Intimate partner stalking. https://www.peaceoverviolence.org/intimate-partner-stalking

Perry, J. (2012). Digital stalking: A guide to technology risks for victims. https://www.womensaid.ie/assets/files/pdf/digital_stalking_guide_v2_nov_2012.pdf

Roberts, K. A. (2005). Associated characteristics of stalking following the termination of romantic relationships. *Applied Psychology in Criminal Justice, 1*, pp. 15–35.

Scott, A., Gavin, J., Sleath, E., & Sheridan, L. (2014). The attribution of responsibility in cases of stalking. *Psychology, Crime & Law, 20*, pp. 705–721. https://doi.org/10.1080/1068316X.2013.854793

Sheridan, L., Arianayagam, J., & Chan, H. C. (2018). Perceptions and experiences of intrusive behavior and stalking within a culture. *Psychology, Crime & Law, 25*, pp. 381–395. https://doi.org/10.1080/1068316X.2018.1529233

Stark, E. (2007). Coercive control: How men entrap women in personal life. *Interpersonal Violence*. Oxford University Press.

Stark, E. (2012). Re-presenting battered women: Coercive control and the defense of liberty. In conference. Violence Against Women: Complex Realities and New Issues in a Changing World, Les Presses de l'Universite du Quebec, Quebec, Canada.

Tassy, F., & Winstead, B. (2014). Relationship and individual characteristics as predictors of unwanted pursuit. *Journal of Family Violence, 29*, pp. 187–195. https://doi.org/10.1007/s10896-013-9573-2

The Charley Project (last updated January 17, 2019). The Charley Project. http://charleyproject.org/case/yessenia-ivette-suarez

Tidy, J. (2019). Stalkerware: The software that spies on your partner. *BBC News*. https://www.bbc.com/news/technology-50166147

Woodlock, D. (2013). Technology-facilitated stalking: Findings and recommendations from the SmartSafe Project. *Domestic Violence Resource Centre Victoria, Collingwood*. https://www.dvrcv.org.au/sites/default/files/SmartSafe_0.pdf

Woodlock, D. (2015). ReCharge: Women's technology safety, legal resources, research and training. *ReCharge Project. Domestic Violence Resource Centre Victoria, Collingwood*. https://www.dvrcv.org.au/sites/default/files/ReCharge_0.pdf

https://doi.org/10.1142/9789811232411_0003

# Chapter 3

# Digital Self-Harm: A Peek into the Mind of an Online Self-Aggressor

John Yu

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*john_yu_from.tp@mha.gov.sg*

## 3.1 Introduction

"You're pathetic and don't deserve to be alive."
"If U don't kill yourself tonight, I'll do it for you."

Insults and threats like the ones above were hurled at a teenager on a popular social media platform by an anonymous user. Dr. Justin W. Patchin, a prominent researcher in the field of cyber-bullying, recounted the incident of a police officer investigating this case of digital harassment [Patchin, 2017]. The culprit's digital trail was eventually traced with the help of the social media company, and the harasser and victim were found to be one and the same person. This unusual account might seem exceptional; yet, researchers and mental health professionals have recently described worrying incidents of students who launched hurtful and harmful comments and contents at themselves on virtual platforms via anonymous personas [Fraga, 2018]. This wilful act of anonymously posting, sending, or sharing of hurtful content about oneself on online platforms has been termed "digital self-harm," and is also known as self-cyberbullying or self-trolling [Patchin & Hinduja, 2017].

The first description of digital self-harm could be attributed to technology researcher Danah Boyd in 2010 when she wrote about it occurring on Formspring, a former social networking site where users can post questions and provide anonymous responses to one another. Boyd posited that there might be three reasons that drive these users to do so. Firstly, it may be a cry for help to attract attention and support from others who would care for them in response to the attacks or insults. Secondly, it may be a way of looking cool because in some schools, receiving criticism is a sign of popularity. Thirdly, it is a mechanism to trigger compliments from friends who will stand up for them and respond with positive words to counteract the negativity [Boyd, 2010]. These attempts may reflect the digitally connected teenager's attempt at coping with one's intra- and inter-personal needs and stresses through social media.

Digital self-harm subsequently took a darker turn when the tragic news of two separate cases of teenage suicide, with elements of self-cyberbullying, were reported. In 2013, 14-year-old Hannah Smith from Leicestershire, England, was found to have anonymously cyberbullied herself with a series of disturbing and hateful messages on social network site ASKfm in the months prior to taking her own life [Davies, 2014]. In Texas, 15-year-old Natalie Natividad, a victim of bullying in school and on social media, was discovered to have posted negative and hurtful statements to herself on a social network mobile app known as After School, before she committed suicide in 2016 [Towner, 2016]. These unfortunate cases have highlighted how digital self-harm could be a novel warning sign of a teenager experiencing a significant amount of emotional distress, or a contributor to negative psychological well-being, or both.

Therefore, the issue of digital self-harm requires greater examination and scrutiny as this may be an emerging, deviant cyber behaviour among digitally embedded teenagers who create online identities and participate in various communities in cyberspace. A greater understanding of why some adolescents would launch verbally and emotionally abusive content at their online persona would also be useful for educators, professionals, and caregivers to be more aware, sympathetic, and able to provide timely and appropriate support. To this end, this article will first provide a key summary of the existing literature on digital self-harm to consolidate what is currently known about it. These empirical findings will provide a

starting point to explore relevant information and possible associations within other related domains, such as the psychology of cyberspace and the cyber-self, bullying, and self-harm. These inferences will then be weaved together in an attempt to provide a probable glimpse into the mind of one who has committed digital self-harm.

## 3.2 Past Studies on Digital Self-Harm

As this is a new phenomenon, there have been limited studies available that examine the underlying reasons fuelling such perplexing behaviours. To date, only three studies have been identified at the time of writing to have conducted surveys on digital self-harm among adolescents and college students in the US and New Zealand [Englander, 2012; Pacheco *et al*., 2019; Patchin & Hinduja, 2017]. These studies have presented important findings on the prevalence of digital self-harm within the adolescent communities, and also discussed the motivations behind digital self-harm, as well as possible correlates that might increase the likelihood or frequency of this behaviour.

### 3.2.1 *Prevalence*

The three studies revealed that the prevalence rate of digital self-harm among teenagers is estimated to be between the range of 6% to 9%. Englander [2012] first reported that 9% out of 617 college students in the US had cyberbullied themselves during high school. A subsequent survey by Patchin and Hinduja [2017] across 5593 students aged 12 to 17 years old revealed that 6.2% had anonymously posted something hurtful about themselves online. Of which, 37.2% said they had done it a few times while 13.2% had done it multiple times. In New Zealand, a 2018 survey of 1,110 teenagers found that 6% had "anonymously posted or shared online mean or harmful content about themselves in the past year." Fifty-seven percent of them had done this a few times while 8% had done so many times [Pacheco *et al*., 2019].

In terms of gender, Englander [2012] reported that more males (13%) than females (8%) engaged in digital self-harm. Likewise, Patchin and Hinduja [2017] reported a smaller but significant difference between males (7.1%) and females (5.3%). Pacheco and colleagues [2019]

reported a similar finding of 7% males and 5% females, although this difference was not statistically significant in their results.

### 3.2.2 *Motivations*

The studies identified the motivations of digital self-harmers by providing pre-identified categories for respondents to select from [Englander, 2012; Pacheco *et al*., 2019], or giving open-ended questions and constructing salient themes from the responses [Patchin & Hinduja, 2017]. In summary, the motivations behind digital self-harm can be categorised into five broad themes as shown in Table 3.1. It is likely that different individuals who engage in digital self-harm would exhibit different permutations of these motives rather than all of them at the same time.

### 3.2.3 *Significant factors associated with digital self-harm*

All three studies reported that teenagers with mental health issues or depressive symptoms were more likely than their peers to have engaged in digital self-harm. Two studies identified the use of drugs and alcohols with digital self-harm behaviour [Englander, 2012; Patchin & Hinduja, 2017]. Other factors that significantly contributed to the engagement of digital self-harm included the presence of one or more physical or behavioural disability, prior participation in deviance, victimisation from school and online bullying,

Table 3.1. Motivations behind digital self-harm.

| | **Motivations underlying digital self-harm** |
|---|---|
| **1** | To prank oneself or others to generate humour or get them into trouble |
| **2** | To measure their level of friendships with others by testing their reactions |
| **3** | To cope with negative emotions, such as sadness or self-hatred |
| **4** | To create an opportunity to publicly display one's resilience in the face of criticism |
| **5** | To indirectly seek help from others |

*Source*: Based on Englander, 2012; Pacheco *et al*., 2019; Patchin & Hinduja, 2017.

engagement in physical self-harm behaviour, and identification as non-heterosexuals [Patchin & Hinduja, 2017; Pacheco *et al*., 2019].

## 3.3 Understanding Digital Self-Harm

The following section presents potential drivers as to why teenagers commit digital self-harm, from a psychological and behavioural perspective. As digital self-harm occurs in cyberspace, it is pivotal to first consider how the infrastructure of cyberspace and the psychology of the online user interact to generate the intention to commit digital self-harm. Since digital self-harm has been said to resemble cyberbullying and physical self-harm [Patchin & Hinduja, 2017], salient theories and perspectives from these domains will be presented to provide possible frames of viewing and understanding why self-cyberbullying occurs online.

### 3.3.1 *Anonymity and disinhibition in cyberspace facilitate digital self-harm*

> "He would log onto one of the four fake accounts he created for self-cyberbullying and send messages in between his parents' laptop and Blackberry phone. One side would be encouraging himself to physically harm his body, 'lmk (let me know) if u need a knife', and the other side would respond, naïve and placid, 'sure hmu (hit me up).'" [Lord, 2018]

The ability for a person to create multiple, unique identities that can interact with one another or others to commit acts of cyber deviance is a product of the digital environment. The nature of electronic communication within the virtual world is characterised by disembodiment as every individual is not constrained to a specific and rigid online persona across multiple networked communities. A single person can create multiple online identities across many online platforms without having to publicly link their actual selves to them, thus enabling anonymity in their ownership and activities. In addition, virtual identities are not only anonymous in nature but 'ephemeral'—they can be created as temporary shells with little need to invest in their

characteristics or persistence [Donath, 1999]. Therefore, virtual identities can be moulded, discarded, and recreated without much inconvenience and commitment.

This virtual anonymity and deindividuation give rise to online disinhibition, where individuals become emboldened to behave differently in cyberspace than how they would in the real world [Suler, 2004]. Therefore, deviant and harmful acts such as cyberbullying, trolling, fraud, and other illegal or unethical activities may be committed by individuals online even if they have not committed any equivalent behaviours in real life. The unusual act of digital self-harm might represent a new cyber behaviour that is driven by the above factors of identity multiplicity, anonymity, and disinhibition.

### 3.3.2 *Digital self-harm as an indirect mean to evaluate one's self-concept through online peer feedback*

Theories of self-concept applied to the real world and cyberspace might be useful to help us understand how adolescents create and manage their identities, as well as present themselves within the online communities. Firstly, the concept of the "looking glass self" purports that self-concept is influenced not only by the individual's internal identities and values, but also of what one perceives how others might view him or her [Cooley, 1902; Sebastian *et al*., 2008]. This implies that one's sense of self is significantly influenced by one's social network and interactions. Teenagers, as part of their identity formation, may be more concerned about how their peers view them and their level of approval towards them as they find it important to fit in with their peers and to be liked [Bellmore & Cillessen, 2006]. These common concerns shared by the youths may affect their sense of self and self-worth during this developmental stage of adolescence [Harter *et al*., 1996].

Jones [2015] further proposed that social media is a modern digital looking glass lens for the contemporary individual, as one's social media presence is public or semi-public and can solicit multiple and immediate feedback from members of one's social community. Therefore, one might be particularly attentive to one's self-presentation on social media and seek to understand how others might feel about themselves, and thereafter exhibit changes to their self-concept based on these perceived results. In the context of digital self-harm, an individual might employ the use of

an anonymous 'other' to instigate responses about oneself from real others to aid in one's inquiry or exploration of the self. Positive replies given in response to negative comments might appear as messages of peer approval and acceptance, thereby creating a positive feeling in the recipient as they elevate one's sense of self-worth.

### 3.3.3 *Digital self-harm as a form of anonymous disclosure of one's negative perception of self*

> "I always seemed like a confident and happy person … But I also wanted people to know that I couldn't always be OK because I was secretly dealing with really bad anxiety. I thought that if I posted that myself out of the blue, people would be like, 'Who are you trying to be? Nobody cares what you think.' I felt like I didn't have a platform unless someone opened the conversation for me. So then I was like, 'Who better to open the conversation than me?'" [Ktena, 2018]

Apart from eliciting feedback from others, cyber platforms provide individuals with a virtual avenue that facilitates self-expression. This implicates another theory of self-concept—the notion of the "true self." One's true self can be described as a reflection of what one intrinsically believes, thinks and feels, which may be different from the actual self which one presents before others as a result of social expectations or constraints [Bargh *et al*., 2002]. The true self can be further delineated into the positive and negative true self. The positive true self refers to the sum of one's personalities, beliefs, and emotions, which are perceived to be more positive in relation to sociocultural norms and values. The negative true self covers the opposite half of the spectrum and is construed by oneself to be less socially desirable. Individuals tend to feel that it is easier to express their negative true self online than in the real world due to elements of anonymity, the lack of normative restraints within cyberspace, a sense of dissociation between their online and offline actions and identities, as well as the presence of other anonymous listeners who can offer support [Hu *et al*., 2017].

However, McKenna and colleagues [2005] found that the presentation of the true self within online communities may be inhibited when there is an offline friend who is able to observe these interactions. This may occur within certain social networking sites where complete anonymity is

not possible and an individual's profile may be known to real-life acquaintances, such as on Facebook or Twitter. This presents as an obstacle to the disclosure of the negative true self, as there is the risk or discomfort of revealing information which are discrepant with one's actual self, and may result in negative social evaluation or disapproval by others [Marriott & Buchanan, 2014]. With regard to digital self-harm, individuals who experience negative emotions may want to express their true selves online, but at the same time find it difficult to articulate them in an online environment where offline peers are present. Therefore, the creation of an anonymous other might facilitate this process by revealing the individual's negative emotional expressions from a third party. This method becomes a safe avenue to express one's negative true self without the perceived fear and judgement of self-disclosure.

### 3.3.4 *The presence and permissibility of the peer norm of cyberbullying may encourage digital self-harm*

> "After this (bullying) happened at school, and online, I became very depressed. I didn't like myself very much. I felt like I deserved to be treated that way, so I thought I would get in on the 'fun'." [Patchin & Hinduja, 2017]
>
> "So people could see that people bully me too and that I could be mean to other people because 'people' were mean to me." [Patchin & Hinduja, 2017]

Since victims of online and school bullying have a higher likelihood of engaging in digital self-harm, it is worth postulating the reasons behind this association from the literature on cyberbullying. Studies have found that being a victim of cyberbullying is often a strong predictor of one turning into a cyber-bully, with outcomes of such bully-victims being more negative than pure bullies, or those who have never been bullied [Hood & Duffy, 2018]. The reason why victims become bullies is often linked with the attributes of cyberspace such as anonymity, reduced social cues, and lack of adult supervision that make retaliatory cyberbullying easier than physical bullying [Antoniadou & Kokkinos, 2015; Mishna *et al*., 2011]. In addition, the normative social influence of an online environment characterised by the presence and permissibility of cyberbullying among peers may cause an individual to accept and adopt such bullying behaviours as a way of conforming to the group norm [Bastiaensens *et al*., 2016; Hinduja & Patchin, 2013].

With the above in mind, digital self-harmers might belong to a unique subgroup of cyber bully-victims who have been impelled by the anonymity of cyberspace and conformity to social norms to engage in cyberbullying of their own selves. We can argue that by doing so, these individuals limit the consequences of their actions as there are no other victims involved. At the same time, they are still able to cyber-bully another persona as a mean to exercise their autonomy and express their belonging to the group norm.

### 3.3.5 *Digital self-harm as a maladaptive form of coping that is self-reinforcing*

> "I wasn't in a good place. I wanted to bash my stupid head through the mirror. Digital self-harm is easier than hurting yourself physically—you don't have to worry about hiding the scars. When you're calling yourself worthless and ugly, it hurts, but it also feels like you're getting everything out your system." [Lord, 2018]

Excluding the motives of creating mischief or humour, the self-derogatory act of cyberbullying oneself ironically aids in producing positively desired outcomes for digital self-harmers, namely coping with negative emotions and achieving the objectives of relationship-testing and support-seeking. This echoes the extant literature on non-fatal self-harm among adolescents which has presented significant findings on the self-reinforcing functions of self-injurious behaviour. For example, self-harmers typically report positive affect regulation in terms of reduction in feelings of depression, anxiety, stress, and anger, during and after the act [Klonsky, 2009; Laye-Gindhu & Schonert-Reichl, 2004].

In their review of 152 studies with first-hand accounts of the reasons for self-harm, Edmonson and colleagues [2016] found several themes of self-harm that can be construed as positive or adaptive experiences for self-harmers. To many, self-injurious behaviours can provide feelings of relief, excitement, comfort, and pleasure. Some participants found self-harm to be a way of self-validation by demonstrating one's strength and toughness, while others reported a sense of self-mastery through regaining a sense of control during and after the act. Lastly, they also found that self-harm was useful to achieve desired interpersonal influence, such as receiving attention, care, and support from others. Interestingly, as self-harm possibly contributes to emotional, interpersonal, and even neurobiological

regulation[a], some have proposed to conceptualise it as a form of behavioural addiction, particularly for people who frequently and repeatedly engage in self-harming behaviours [Blasco-Fontecilla *et al*., 2016].

> "A bunch of people I was friends with online would say, 'Oh don't listen to the hate.' And then that kind of gave me this satisfaction, like when someone likes your post on Instagram. It gave me that same feeling. I think it can become a bit of an addiction." [Ktena, 2018]

Therefore, there is a similarity between digital and physical self-harm as both are maladaptive forms of coping that yield personally desirable outcomes at the cost of self-inflicted suffering and pain. The paradoxical derision and demeaning of oneself is a temporary pain to endure for the positive after-effects that can come as a cathartic release and regulation of emotion, as well as an outpouring of social support and validation. Digital self-harm can be considered a virtual variant of physical self-harm in which no physiological pain has to be experienced but positive psychological benefits can still be yielded. This might explain why Patchin & Hinduja [2017] found that people who participated in offline self-harm were significantly more likely to be involved in digital self-harm. Even so, just as habitual self-harm is maladaptive and can lead to worst outcomes, digital self-harm should also be viewed with concern, particularly for individuals who resort to such behaviour as their primary coping method to tide through emotionally and situationally stressful periods.

### 3.3.6 *Online trend of receiving positive affirmation after public displays of resilience may promote digital self-harm*

> "Being called a slut or fat or whatever, it made me feel seen. It sounds weird but celebrities and popular people are always the ones who receive hate. That's when you know you're cool, when people can be arsed telling you to die … I wrote an open letter about how the troll's comments made me feel. Everyone in my class was really supportive, calling me brave and thanking me for having the courage to talk about my anxiety so frankly." [Lord, 2018]

---

[a]Blasco-Fontecilla and colleagues [2016] suggests that just like other behavioural addictions, the opioid, dopaminergic, and overactivation of stress systems may be implicated in repeated self-harming behaviours.

Some adolescents commit digital self-harm as a show or proof of their toughness and resilience against online hatred. While the cultivation of youth resilience is an important protective factor against cyberbullying victimisation [Hinduja & Patchin, 2017], this concept takes on a different form in digital self-harm as both the aggressor and resilient victim are the same person. Lord [2018] suggested that digital self-harm is influenced by the prevailing internet culture of how renowned, public personalities are praised and applauded by their followers whenever they publicly respond to online criticisms in positive or stoic manners. Perhaps online attention is generated towards public expressions of resilience as they emphasise inspiring and encouraging narratives that depict how individuals overcome their adversaries or adversities. Such online content that generate positive emotions tend to be more popular or viral, and are more likely to be socially transmitted across a larger network [Berger & Milkman, 2012]. As a result of frequent exposure to such incidents, individuals might be socially influenced to think that the phenomenon of being cyberbullied and thereafter proving one's resilience is a socially approved trend or behaviour by online communities. Digital self-harm can function as an anonymous method to initiate this process of self-attack and self-defence in a controlled manner, in order to achieve the goals of drawing attention to oneself and receiving boosts to one's self-esteem and validation through the influx of supportive comments.

## 3.4 Responding to Digital Self-Harm

### 3.4.1 *Social media and tech companies*

Social network and digital platforms that are widely used by adolescents can play a key role in safeguarding, detecting, managing, and preventing digital self-harm within their user communities. One positive example is ASKfm, a social networking site that is popular for registered or anonymous posting of questions and answers, and the various measures they have implemented to address digital self-harm. Public education in the form of web articles have been published in collaboration with mental health professionals to educate their users on what self-messaging, a more neutral term for digital self-harm, constitutes and how they should respond to it in a healthier manner [ASKfm, n.d.]. Technological measures are also in place, with ASKfm detecting and removing self-hate messages and directing these users to relevant support and resources by

partnering with Koko, an online emotional distress management and response service [Jones, 2018]. This holistic approach covers both content detection and management, as well as the provision of psychoeducation to all users and targeted support to at-risk individuals who display warning behaviours. This may represent a good baseline that social networking companies could adopt in their policies and measures to combat hurtful cyber behaviours within digital platforms.

### 3.4.2 *School and mental health professionals*

Schools, youth social services, and mental health professionals who engage in real-life contact with teenagers presenting with emotional or behavioural problems can play a significant role in detecting whether digital self-harming behaviours might also be present. Raising awareness on the symptoms and functions of this novel cyber behaviour can help them to ask better questions to identify whether an adolescent might be engaging in digital self-harm. With reference to treatment approaches suggested by Beard [2011] on problematic internet use, counsellors or therapists can adopt a biopsychosocial model of assessment to delineate the motivations and functions behind this behaviour. This can include understanding if there are any biological symptoms, psychological factors, and social dynamics that influence the adoption and maintenance of acts of digital self-harm. For example, a teenager who is a victim of bullying at school might engage in digital self-harm to achieve a sense of power and control, whereas a teenager who self-mutilates might cyberbully themselves as a cry for help. Distilling this knowledge would aid professionals in formulating a more specific and effective intervention plan for their teenage clients that includes the management of digital self-harm. Furthermore, existing broad-based or targeted programmes in schools or centres can consider including a segment on digital self-harm education and prevention to more comprehensively address a larger spectrum of deviant cyber behaviours that teenagers might engage in.

### 3.4.3 *Parents at home*

At home, parents have an important role in keeping watch and supporting their children. Studies have shown that parent-child communication, parental monitoring, and the quality of the parent-child relationship are

important factors that can reduce the incidence and impact of cyberbullying and self-harm among adolescents [Buelga *et al.*, 2016; Buelga *et al*., 2017; Klemera *et al*., 2017]. Concerning digital self-harm, Toni Birdsong, a family safety evangelist at McAfee, recommends that parents can listen and pay more attention to their children's wellbeing and behaviours, as well as adopt a non-judgemental approach when their children confide in them about their emotional issues. They can also observe their children's social interactions on social media, read the comments to their posts, and talk to their children about negative or threatening ones. If digital self-harm is indeed present, parents can better decide and guide their children to seek further help or support if required [Birdsong, 2020]. In essence, nurturing close families and adopting positive family-based approaches might be helpful to detect, manage, and reduce digital self-harm among adolescents.

### 3.4.4 *Law enforcement and investigators*

From a law enforcement perspective, digital self-harm can be viewed as a novel variety of cyber deviance that police officers might encounter, investigate, and manage. This method of fabricating threats or attacks on oneself is not new, as it resonates with a modus operandi of staging self-inflicted injuries and attributing them to an innocent party to be the scapegoat for personal or malicious motives. An example would be an individual who files a false rape claim against another and even fabricates evidence of the alleged assault to gain attention or sympathy from others [Chancellor & Graham, 2016]. Therefore, digital forensics remain a relevant and pivotal tool in any investigative work that involves tracing the activities of a suspect in cyberspace and his whereabouts. Investigators who are handling incidents or allegations of excessive cyber harassment could remain open to even the slightest possibility that the victims themselves might be the perpetrators, and ensure that their investigations are thoroughly conducted. If there are reasonable suspicions that arise, underlying motives and agenda are likely to be present and should be further investigated upon. Digital self-harmers who commit the act out of genuine emotional or psychological turmoil can then be referred to proper channels of mental health support, while those who do so out of vicious or even criminal motives can be taken to task.

## 3.5 Conclusion

Digital self-harm is a curious and concerning behaviour that has emerged among adolescents in cyberspace. While this phenomenon came to light a decade ago, there are still only a handful of studies and empirical information that we have on this deviant behaviour. This article has presented some key findings in the literature and attempted to dive deeper into the motivations behind digital self-harm through various lenses—the self and cyberspace, bullying, and self-harm. Digital self-harm is permitted by the powers of anonymity, identity creation, and manipulation that individuals possess online. Depending on the individual, digital self-harm can be a process that is useful for social validation, influence, and conformity within one's networked community, or a functional coping mechanism that can regulate negative emotions by receiving positive attention and support from online others. It has also been viewed as a worrying behaviour particularly for teenagers who rely on it as a maladaptive coping strategy. Therefore, addressing this problem may require concerted efforts from technology companies and those who are physically close to these individuals, to understand them better and help them find healthier means to cope with their stressors and emotions.

As this chapter is largely exploratory and descriptive rather than empirical, it has several limitations that should be kept in mind. Many of the claims linking aspects of digital self-harm with psychological concepts from relevant domains like bullying and self-harm require further studies to validate their associations. After all, it has yet to be ascertained whether digital self-harm should be classified as a subset of an existing deviant behaviour such as cyberbullying or self-aggression, or a completely different category of cyber behaviour on its own. However, this chapter is a humble attempt to understand the reasons underlying the motives of adolescents who commit this curious cyber behaviour and to spark further questions and research. As current studies have only been conducted among high school and college students, it would be interesting to identify whether digital self-harm might be present in adult populations and how these behaviours and motivations might differ or overlap given the differences in demographics.

## 3.6 Acknowledgement

The views expressed in this chapter are the author's alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 3.7 References

ASKfm. (n.d.). *Tackling the Self-Messaging Issue 2017*. ASKfm Safety Centre. https://safety.ask.fm/tackling-self-messaging-issue/

Bargh, J. A., McKenna, K. Y. A., & Fitzsimons, G. M. (2002). Can you see the real me? Activation and expression of the true self on the internet. *Journal of Social Issues, 58*(1), pp. 33–48.

Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2016). From normative influence to social pressure: How relevant others affect whether bystanders join in cyberbullying. *Social Development, 25*(1), pp. 193–211.

Beard, K. W. (2011). Working with adolescents addicted to the Internet. In K. S. Young & C. N. de Abreu (Eds.), *Internet addiction: A handbook and guide to evaluation and treatment* (pp. 173–189). John Wiley & Sons Inc. http://www.ssu.ac.ir/cms/fileadmin/user_upload/Moavenatha/MBehdashti/ravan/pdf/faaliyatha/pptfiles/INTERNET_ADDICTION.pdf#page=193

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), pp. 192–205.

Bellmore, A. D., & Cillesen, A. H. N. (2006). Reciprocal influences of victimization, perceived social preference, and self-concept in adolescence. *Self and Identity, 5*(3), pp. 209–229.

Birdsong, T. (2018, May 18). *Study: Digital Self-Harm Among Teens Real; Here's What Parents Need to Know*. McAfee. https://www.mcafee.com/blogs/consumer/family-safety/study-digital-self-harm-among-teens-real-heres-what-parents-need-to-know/?utm_campaign=Consumer&utm_source=facebook&utm_medium=spredfast&utm_content=

Blasco-Fontecilla, H., Fernández-Fernández, R., Colino, L., Fajardo, L., Perteguer-Barrio, R., & de Leon, J. (2016). The addictive model of self-harming (non-suicidal and suicidal) behavior. *Frontiers in Psychiatry, 7*(8), pp. 1–7.

Boyd, D. (2010, December 7). *Digital self-harm and other acts of self-harassment*. Danah Boyd Apophenia. http://www.zephoria.org/thoughts/archives/2010/12/07/digital-self-harm-and-other-acts-of-self-harassment.html

Buelga, S., Martínez-Ferrer, B., & Musitu, G. (2016). *Family relationships and cyberbullying*. In Navarro R., Yubero S., & Larrañaga E. (Eds.), Cyberbullying Across the Globe. Springer.

Buelga, S., Martínez-Ferrer, B., & Cava, M. (2017). Differences in family climate and family communication among cyberbullies, cybervictims, and cyber bully–victims in adolescents. *Computers in Human Behaviors, 76*, pp. 164–173.

Chancellor, A. S., & Graham, G. D. (2016). *Crime Scene Staging: Investigating Suspect Misdirection of the Crime Scene*. (Charles C Thomas Publishing Ltd).

Cooley, C. H. (1902). *Human nature and the social order*. Scribner's Sons.

Davies, C. (2014, May 6). *Hannah Smith wrote 'vile' posts to herself before suicide, say police*. The Guardian. https://www.theguardian.com/uk-news/2014/may/06/hannah-smith-suicide-teenager-cyber-bullying-inquests

Donath, J. S. (1999). Identity and deception in the virtual community. In P. Kollock & M. Smith (Eds.), *Communities in cyberspace.* (Routledge) pp. 29–59.

Edmonson, A. J., Brennan, C. A., & House, A. O. (2016). Non-suicidal reasons for self-harm: A systematic review of self-reported accounts. *Journal of Affective Disorders, 191*, pp. 109–117.

Englander, E. (2012). *Digital Self-Harm: Frequency, Type, Motivations, and Outcomes*. MARC Research Reports. http://vc.bridgew.edu/marc_reports/5

Fraga, J. (2018, April 21). *When teens cyberbully themselves*. National Public Radio. https://www.npr.org/sections/health-shots/2018/04/21/604073315/when-teens-cyberbully-themselves

Harter, S., Stocker, C., & Robinson, N. S. (1996). The perceived directionality of the link between approval and self-worth: The liabilities of a looking glass self-orientation among young adolescents. *Journal of Research on Adolescence, 6*(3), pp. 285–308.

Hinduja, S., & Patchin, J. W. (2013). Social influences on cyberbullying behaviors among middle and high school students. *Journal of Youth and Adolescence, 42*(5), pp. 711–722.

Hinduja, S., & Patchin, J. W. (2017). Cultivating youth resilience to prevent bullying and cyberbullying victimization. *Child Abuse & Neglect, 73*, pp. 51–62.

Hood, M., & Duffy, A. L. (2018). Understanding the relationship between cyber-victimisation and cyber-bullying on Social Network Sites: The role of moderating factors. *Personality and Individual Differences, 133*, pp. 103–108.

Hu, C., Kumar, S., Huang, J., & Ratnavelu, K. (2017). Disinhibition of negative true self for identity reconstructions in cyberspace: Advancing self-discrepancy theory for virtual setting. *PLoS ONE, 12*(4): e0175623.

Jones. J. (2015). The looking glass lens: Self-concept changes due to social media practices. *The Journal of Social Media in Society, 4*(1), pp. 100–125.

Jones, S. (2018, May 30). Teens are cyberbullying themselves. Why? Education Week. https://www.edweek.org/ew/articles/2018/05/30/teens-are-cyberbullying-themselves-why.html

Klemera, E., Brooks, F. M., Chester, K. L., Magnusson, J., & Spencer, N. (2017). Self-harm in adolescence: protective health assets in the family, school and community. *International Journal of Public Health, 62*, pp. 631–638.

Klonsky, E. D. (2009). The functions of self-injury in young adults who cut themselves: clarifying the evidence for affect regulation. *Psychiatry Res*, *166*(2–3), pp. 260–268.

Ktena, N. (2018). *These teens secretly trolled themselves online*. BBC. https://www.bbc.co.uk/bbcthree/article/05e9991d-4713-4ad4-b9af-eecd47d7dfd7

Laye-Gindhu, A., & Schonert-Reichl, K. A. (2005). Nonsuicidal self-Harm among community adolescents: Understanding the "Whats" and "Whys" of self-harm. *Journal of Youth and Adolescence*, *34*(5), pp. 447–457.

Lord, A. (2018, June 27). *Self harm can be digital too*. Vice i-D. https://i-d.vice.com/en_uk/article/kzkvxz/self-harm-can-be-digital-too

Marriott, T. C., & Buchanan T. (2014). The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline. *Computers in Human Behavior*, *32*, pp. 171–177.

McKenna, K. Y. A., Buffardi, L., & Seidman, G. (2005). Self-presentation to friends and strangers online. In K-H. Renner, A. Schutz, & F. Machilek (Eds.), *Internet and Personality* (Hogrefe and Huber), pp. 175–188.

Pacheco E., Melhuish, N., & Fiske, J. (2019, May 1). *Digital self-harm: Prevalence, motivations and outcomes for teens who cyberbully themselves*. Netsafe. https://www.netsafe.org.nz/wp-content/uploads/2019/05/Digital-self-harm-report-2019.pdf

Patchin, J. W. (2017, October 3). *Digital self-harm: The hidden side of adolescent online aggression.* Cyberbullying Research Center. https://cyberbullying.org/digital-self-harm

Patchin, J. W., & Hinduja, S. (2017). Digital self-harm among adolescents. *Journal of Adolescent Health, 61*(6), pp. 761–766.

Sebastian, C., Burnett, St., & Blakemore, S. (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences, 12*(11), pp. 441–446.

Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior, 7*(3), pp. 321–326.

Towner, M. (2016, November 8). *Texas girl, 15, commits suicide after she was bullied in school and online, family claims*. DailyMail Online. https://www.dailymail.co.uk/news/article-3899914/Girl-15-commits-suicide-allegedly-bullied-school-online.html

This page intentionally left blank

# Section B

# Sexual and Deviant Behaviours Online

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0004

# Chapter 4

# Cyber Sexual Deviance: Delving into Image-Based Sexual Abuse

Vivian Seah

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Vivian_Seah@mha.gov.sg*

## 4.1 Introduction to Cyber Sexual Deviant Behaviour

In 2011, Holly Jacobs received an email that changed her life. It was from a stranger who tipped her off about a website that contained sexually explicit images of her. Thereafter, threatening emails came along—emails that threatened to send those images to her employer, co-workers, and the general public [Jacobs, 2013]. Three days later, the threats came true, and her compromising photos and personal details (i.e., her name, phone number, and email address) could be found on more than 200 websites, and the perpetrator was none other than her ex-boyfriend [Miller, 2013]. She described her ordeal in a *Thought Catalog* article, about how she received countless harassment emails from strangers—telling her how much they were enjoying her photos, some of whom even attached nudes of themselves [Jacobs, 2013]. Even after Jacobs had spent time and effort to take down her compromising photos from the websites, they would resurface on more websites a few weeks later. She has since changed her

email address and phone number, removed herself from social media, changed jobs, and even changed her name legally [Miller, 2013]. The silver lining of Jacobs' story is how she used her personal experience to start the "End Revenge Porn" online campaign in 2012 to help others in a similar plight. Through the campaign, she launched a website to advocate for the criminalisation of non-consensual pornography and provide informational support for other victims alike.

The Internet has transformed various aspects of our lifestyles, including education, entertainment, businesses, and in particular, communication [Durkin *et al*., 2006]. The advances in digital communications technologies have altered the ways we interact and connect with one another. Worryingly, it has led to digital technologies being increasingly used as tools of harassment, abuse, violence, and sexual deviance (i.e., sexual habits and behaviour that go against social norms) [Powell *et al*., 2018]. For instance, Dir and Cyders [2015] found that approximately 18% to 68% of young adults (18 to 24 years old) internationally make use of technology to exchange explicit messages and images. As cyberspace enables individuals with similar deviant interests around the globe to gather and interact with one another, sexually deviant individuals could seek affirmation and reinforcement for their norm-opposing actions through the Internet [Durkin, 2001]. For instance, pathological sexual practices and deviant acts such as bondage-domination, bestiality, and incest can be found through internet pornography [Quinn & Forsyth, 2005]. While these deviant behaviours were disapproved of in the past, they are now being featured or even accepted in the cybersphere.

However, it is noteworthy that not all cyber sexual deviant behaviours are criminalised. According to Griffiths [2000], sexually-related Internet deviant behaviour that are typically recognised as crimes can be categorised into (1) displaying, downloading, and/or distributing of illegal sexually-related material and (2) using the Internet to sexually procure and/or intimidate someone. Some examples of cyber sexual deviant crimes include but are not limited to: cyberstalking, online sexual harassment, paedophilic grooming of child victims, and Image-Based Sexual Abuse (IBSA) [Griffiths, 2000]. With regard to legislation on IBSA, recent amendments to Singapore's Penal Code criminalises the following: "(i) Making, distribution, possession of, and access to, voyeuristic recordings or intimate images and (ii) Distribution of or threat to distribute intimate images or recordings" [Ministry of Law, 2019].

Non-consensual exposure of one's genitals in cyberspace (i.e., cyber-flashing) is also an offence in Singapore [Ministry of Law, 2019].

Holly Jacobs, whose example was highlighted at the start of the chapter, is one of the many on the receiving end of cyber sexual deviance, more specifically, Image-Based Sexual Abuse (IBSA). This chapter will delve into the dark side of the cyber world—on how the internet is used as a tool to perpetrate cyber sexual deviant behaviour. In particular, building on past research on IBSA from sociological and feminist-criminological angles, this chapter seeks to examine the prevalence, underlying motives, and impact of IBSA. Possible mitigation strategies for the issue of IBSA will also be discussed.

### 4.1.1 *What is image-based sexual abuse?*

Image-Based Sexual Abuse (IBSA) is a term also known by some as non-consensual pornography or revenge pornography [Powell *et al*., 2018]. "Revenge pornography" is described as the non-consensual distribution of sexually explicit materials by ex-partners to seek revenge after the relationship ends [Powell *et al*., 2019]. Researchers and academics argue that the concept of "revenge pornography" is too narrow and does not adequately address the different nature, motivations, and extent of harms of the various cyber-facilitated sexual offences involving images or videos [McGlynn *et al*., 2017]. Therefore, this chapter will adopt a broader definition of IBSA that encompasses a myriad of technology-facilitated sexual crimes, beyond just revenge pornography.

In particular, IBSA encompasses the following: (i) *taking or creating* of non-consensual sexual images or videos; (ii) *sharing or distribution* of non-consensual sexual images or videos; and/or (iii) *threatening* to take, share or create non-consensual sexual images or videos of another party [McGlynn *et al*., 2019]. This definition of IBSA adopted would cover a broader set of technology-enabled sexual crimes, including: (i) relationship retribution; (ii) sextortion; (iii) sexual voyeurism; (iv) sexploitation; and (v) filming of sexual assault [refer to Table 4.1 for a summary of details; Powell & Henry, 2017; Powell *et al*., 2018]. The five typologies are differentiated by their definition, modus operandi, and underlying motivation as highlighted in Table 4.1. The underlying motivations of the various IBSA typologies will be further elaborated upon later in this chapter.

Table 4.1. Key typologies of Image-Based Sexual Abuse.

| S/N | Typology | Definition/ Modus Operandi | Key motivation | Example |
|---|---|---|---|---|
| 1 | Relationship Retribution | Defined as the act of using shared intimate images/ videos as a tool by vindictive ex-lovers to humiliate, embarrass or "seek revenge" against their ex-partners [Henry & Powell, 2017]. | Alleviating grievance and revenge; Social reinforcement | A 21-year-old Singaporean man sent 20 previously consensually shared nude images, along with the name and school details of his ex-girlfriend to a Tumblr site administrator after he had found out that she was communicating with another man and he was enraged [Chong, 2017]. |
| 2 | Sextortion | Defined as the act of threatening to disseminate sexually explicit images/videos of a person in order to coerce him/her into fulfilling certain demands [Cyber Civil Rights Initiative, 2020]. Sextortion has been perpetrated against intimate partners, friends, acquaintances, or even strangers [Henry & Powell, 2017]. | Financial gain | In 2016, a 44-year-old man in Singapore obtained nude photographs and the contact number of a 25-year-old woman after her ex-boyfriend posted a message online, inviting people to contact him privately for those nude photographs. With these, the man contacted the victim and threatened her with the options of either using money or her body to "redeem" the explicit images he had of her [Alkhatib, 2018]. |
| 3 | Sexual Voyeurism | Defined as the creation or distribution of images/ videos taken without the knowledge of the victim, often as a form of sexual gratification [Powell & Henry, 2017]. | Sexual gratification; Social reinforcement | Examples of sexual voyeurism acts include:<br>• Covert filming of consensual sexual acts<br>• Upskirting/Down-blousing—taking photographs/videos up unsuspecting women's skirts/dresses or down their blouses with the use of camera-equipped devices [Nolan & Maguire, 2016]<br>• Manipulating victims' computers (e.g., via hacking, phishing, malware) to obtain and disseminate sexually explicit materials without their consent [McGlynn *et al*., 2017]. |
| 4 | Sexploitation | Defined as the cyber commercial trade of non-consensual sexual materials with the primary goal of financial exploitation [Powell & Henry, 2017]. | Financial gain | A victim had a compilation of her naked photos posted on eBay auction by her ex-boyfriend who wanted to make profits with her nude photos [Bates, 2017]. |
| 5 | Filming of Sexual Assault | Defined as the creation and/or distribution of sexual assault images or videos recorded by perpetrators and/or bystanders of the rape [Powell & Henry, 2017]. | Sexual gratification; Sense of power | A US woman was drugged and raped by her husband and another man, with the whole incident filmed and sent to her colleagues after they divorced [Bates, 2017]. |

## 4.2 The Prevalence of Image-Based Sexual Abuse

Image-Based Sexual Abuse (IBSA) is a growing concern, with increasing prevalence and victimisation in various countries. Recent studies have found that IBSA appears to be an emerging "trend" of concern [McAfee, 2013; Thompson & Morrison, 2013; Powell & Henry, 2015; Henry, *et al*., 2017; Powell *et al*., 2019]. In the US, a study by McAfee [2013] found that one in 10 participants (n = 1,182) were threatened by their ex-partners to have their sexually explicit photos posted online, and such threats were carried out 60% of the time. In another study conducted with 795 American men, it was found that 16% of them had shared with others a sexually suggestive image of another person without the party's prior consent [Thompson & Morrison, 2013].

In Australia, a study on 2956 respondents in 2014 found that one in 10 Australians regardless of gender, reported experiences of their nude or semi-nude image taken by others without their consent [Powell & Henry, 2015]. In a similar survey conducted in 2016, the percentage of IBSA victimisation in Australia (n = 4,274) doubled from 2014—i.e., one in five Australians between ages 16 to 49 reported experiencing at least one form of IBSA in 2016 [Henry *et al*., 2017].

In Singapore, there has been an increase in help-seeking cases from victims of technology-facilitated sexual violence. The Sexual Assault Care Centre (SACC) of the Association of Women for Action and Research (AWARE) reported 46 cases (out of 338 total cases) in year 2016, 99 cases (out of 515 total cases) in 2017, and 124 cases (out of 808 total cases) in 2018 [AWARE, 2019]. In 2019, Singapore witnessed an increase in IBSA cases that sparked public outcry. Voyeurism cases within universities in Singapore gained media attention after a local undergraduate, Monica Baey, took to Instagram to share her experience of being filmed by a fellow male student in the bathroom of her school's residence hall in April 2019 [Goh, 2019]. In October 2019, a few Telegram chat groups with Singapore users sharing non-consensual pornographic materials were discovered; one of the chat groups named "SG Nasi Lemak" had more than 44,000 members ["4 arrested for allegedly circulating obscene materials," 2019]. Following these incidents, IBSA began to gain traction and public attention, with calls for change in Singapore's stance towards IBSA [Sim, 2019].

In addition, there has been worldwide concern regarding the rise in IBSA during the Coronavirus Disease (COVID-19) lockdowns in various countries since March 2020—as the increase in time spent at home during

lockdown has led to an increase in online activities and perpetration of cybercrimes such as IBSA [Davies, 2020]. The increase in cases of IBSA follows the trend of increased porn site traffic during COVID-19 lockdown, for instance, it is reported that there was a 57% increase in site traffic for pornographic website "Pornhub" in Italy, a 39% rise in France, and a 61% rise in Spain, after Pornhub had promoted their website by providing free Premium access to residents in these countries during the lockdown in March [Grant, 2020]. It has been a worrying trend—with UK's Revenge Porn Helpline receiving two times more calls for help in April 2020, as compared to April 2019 [Davies, 2020]. Therefore, it is necessary to delve into understanding IBSA as an issue, in order to explore effective mitigation strategies for it.

## 4.3 Driving Factors for Perpetrators of Image-Based Sexual Abuse

In view of the different typologies and the prevalence of IBSA highlighted, the obvious question will be: Why do people commit IBSA? To that, the "Rational Choice Theory" (RCT) can be adapted as a general theory of crime to explore and understand why people commit IBSA—when the perceived benefits outweigh the costs of IBSA. When applied to deviant and criminal behaviour, the RCT suggests that a rational offender weighs the risks, costs, and benefits of a crime, and commits the act only when the perceived benefits outweigh the risks and costs [Becker, 1968; Loughran *et al*., 2016].

According to Becker [1968], one tends to be deterred from committing an offence when the perceived certainty and severity of the punishment (i.e., likelihood of being jailed or fined, etc.) outweigh the perceived benefits (i.e., monetary, intrinsic, or social benefits, etc.) of partaking in the crime. As detailed in Table 4.1, the different typologies of IBSA may stem from differing motivations. This section seeks to highlight and discuss the motivations, driving factors, and benefits perceived by perpetrators of IBSA—specifically, alleviating grievance and revenge, financial gain, social reinforcement, sexual gratification, and achieving a sense of power. Perpetrators of IBSA may perceive the aforementioned "benefits" of IBSA to outweigh the relatively low likelihood of apprehension and the low severity of legal consequences for IBSA.

### 4.3.1 *Perceived benefit of IBSA #1: Alleviating grievance and revenge*

The alleviation of one's grievance and revenge is seen as a motive to commit IBSA, especially for relationship retribution typology, where people are driven to share revenge pornography of their ex-partner to harm and "get even" with their ex-partner [Sirianni & Vishwanath, 2016]. Sirianni and Vishwanath [2016] also found that individuals who were highly provoked by their ex-partner were more likely to share sexually explicit materials of them, with hopes of damaging or harming their ex-partner in some way, such as making them feel regret, humiliation, and suffering. The more individuals perceive that the act of revenge is able to bring desired negative changes to their ex-partner's attitudes, feelings, behaviour, or lifestyle, the more "benefits" the act of IBSA seem to bring to the perpetrator, such as the reparation of one's ego, self-esteem, and power [Boon *et al*., 2011]. This may be done by privately sharing the victims' intimate materials with others via online chats or email, and/or publicly on social media, forums, blogs, or pornographic sites without the consent of their ex-partner [Henry & Powell, 2017].

In a study conducted by Uhl and colleagues [2018] that examined the reasons for posting non-consensual sexual materials on websites, the title and description of sexual images were analysed and the results showed that the most common information include hateful insults towards the victim (e.g., "cheating, man-eating bitch") and description that identifies the victim as an ex-partner (e.g., "former ex-girlfriend"). This shows that the online posting of non-consensual sexual materials is often attributed to some form of grievance against one's ex-partner—with the goal being to exact revenge and alleviate one's grievance against the victim.

### 4.3.2 *Perceived benefit of IBSA #2: Financial gain*

Another motivation for IBSA perpetrators could be financial gain. Inferred from various examples, IBSA perpetrators tend to use three main ways to make financial gain out of IBSA: (1) through advertising means on non-consensual pornography websites; (2) direct online sales of non-consensual pornography; and (3) extortion of money from the victims in exchange for the removal of the victim's sexually explicit photos online.

#### 4.3.2.1 *Financial gain through advertising means*

As detailed in Table 4.1, non-consensual sexual materials may be exploited for financial gain by perpetrators through sexploitation. In a well-known case that took place in the US, a man named Hunter Moore created a website, isanyoneup.com, in 2010 to feature sexually explicit materials of both men and women, along with derogatory comments and their personal details such as their full name, profession, link to their social media profile, and residential address [Morris, 2012]. The website received a high traffic count of up to 350,000 site visitors a day, and Moore was earning approximately US$13,000 a month through advertising revenue from the website before it shut down in 2012 [Greenhouse, 2014].

With just a Google search, many websites dedicated to non-consensual pornography can be found—for instance, Expic.net, Watchmyexgf.com, the now-defunct MyEx.com, and many other pornography websites with ex-girlfriend pornography as one of the genres. This shows that non-consensual pornography may be highly sought after as a new genre of pornography for sexual gratification. Therefore, IBSA perpetrators who are opportunists may have created standalone non-consensual pornography websites to generate financial gain through advertising means.

#### 4.3.2.2 *Financial gain through direct sales of non-consensual sexual materials*

In another form of sexploitation with a different modus operandi, perpetrators of IBSA gain financial profit through the sales of non-consensual pornography. This can be perpetrated via various online means, such as online forums, websites, and online messaging applications. In Singapore, a 20-year-old man was arrested for selling nude images of his female victims on the messaging application, WeChat [Lum, 2018].

Perpetrators of IBSA may even exploit deviant online community platforms to "publicise" their non-consensual pornography collection. This was observed in one of the Telegram public groups that shared non-consensual pornographic materials, whereby one member had publicised the sales of "1tb of personal collection of girls I know" ["Leaked sex tapes and child porn," 2019].

#### 4.3.2.3 *Financial gain through extortion of money from victims*

Victims of sextortion may be extorted in exchange for having their sexual images removed online or deleted by the perpetrator. For instance, the website "Is Anybody Down?" that hosted non-consensual pornographic materials between 2011 to 2013 profited from women who paid for the removal of their sexual images from the website [Salter & Crofts, 2015].

Although the financial gains from IBSA may seem lucrative to some, other perpetrators may still be motivated to perpetrate IBSA due to other reasons, such as social benefits.

### 4.3.3 *Perceived benefit of IBSA #3: Gaining social reinforcement*

Some perpetrators of IBSA may have committed the act (i.e., relationship retribution and sexual voyeurism) to gain social reinforcement, mainly through compliments and acceptance from others on online platforms such as pornographic forums, social media platforms, and online messaging applications. One example of such online communities would be the "SG Nasi Lemak" Telegram group that was busted in Singapore for sharing non-consensual sexually explicit materials among more than 44,000 members ["4 arrested for allegedly circulating obscene materials," 2019]. Similar chat groups were also discovered, and it was found that those who had frequently shared sexually explicit materials were held in high regard by others in the chat and given "premium membership" that entailed an invitation to another "exclusive chat group" ["Leaked sex tapes and child porn," 2019].

As a result of such social reinforcement, perpetrators may continue or even increase the frequency of IBSA. This is in line with the psychological concept of "operant conditioning," which states that when a behaviour is rewarded, there is an increased likelihood of one engaging in the same behaviour in the future [Skinner, 1938]. In a wicked cycle, the online audience of the sexually explicit materials posted by the IBSA perpetrators are provided with sexual gratification from viewing and masturbating to those images, and this enjoyment serves as a reward for them to continue providing social reinforcement to the IBSA perpetrators due to operant conditioning [Skinner, 1938].

### 4.3.4 *Perceived benefit of IBSA #4: Sexual gratification*

Similar to voyeurs who masturbate to their voyeuristic memory, images and recordings obtained through sexual voyeurism and the filming of sexual assault tend to be used by the perpetrators to gain sexual gratification (i.e., to masturbate to) [Holmes & Holmes, 2008]. In a 2014 high-profile case of sexual voyeurism via the manipulation of victims' computers, the Apple accounts of various famous Hollywood stars were hacked, and their intimate photos were compromised and widely disseminated on various online forums and websites [McCoy, 2014]. Although the original photos were taken down, those sexually explicit images were subsequently reposted on various websites as pornographic material for sexual gratification. In Singapore, there were a series of sexual voyeurism cases in 2019, whereby perpetrators of various age groups (i.e., school-going teenager, middle-aged lecturer, and retired elderly) took upskirt videos and images of women in public [Alkhatib, 2019a; Alkhatib, 2019b; Lam, 2019]. These sexually explicit images and videos obtained through sexual voyeurism were subsequently used to fulfil the perpetrators' sexual needs.

### 4.3.5 *Perceived benefit of IBSA #5: Sense of power*

Besides the aforementioned driving factors of IBSA, research suggests that perpetrators of IBSA tend to perceive that they hold a certain degree of power over the victim and that they enjoy this sense of power [Bates, 2017]. Ex-partners may feel a loss of control and power when their relationship with the victim turned sour, which may drive them to commit IBSA to regain their sense of power by controlling how others view the victim [Boon *et al*., 2011]. For perpetrators who are not intimately related to the victim, the act of invading someone's privacy tends to give the perpetrator a sense of control over the victim and the situation, fulfilling their need for power [Calvert & Brown, 2000]. Therefore, the perpetrators of different typologies of IBSA seem to be motivated by a commonly perceived sense of power, and this may be especially true for perpetrators who film sexual assaults in order to revisit their crime. Just as perpetrators of intimate partner violence gain power through coercive control, the individual who films a sexual assault may enjoy the "reward power" (i.e., one's ability to give or take away things from the victim) of domination and control over the degree of suffering he/she is able to inflict on the victim when deciding whether or not to disseminate the film [Dutton & Goodman, 2005; O'Hara, 2019].

### 4.3.6 *Perceived costs of IBSA*

In recent years, laws have been implemented in the US, the UK, Australia, and Singapore to tackle IBSA in the cybersphere and beyond [Su *et al*., 2019; "Revenge porn: What is the law in the UK?," 2019; Wahlquist, 2019; Singapore Legal Advice, 2020]. The magnitude of the punishment is country and state dependent, and case-specific, with jail terms from below one year to up to five years [Nigam, 2018]. As IBSA is not recognised as a sexual crime, sentencing of perpetrators are deemed more lenient—for instance, the founder of revenge pornography website IsAnyoneUp.com, Hunter Moore, was sentenced to a mere two and a half years imprisonment despite the harm he had brought to many victims worldwide through his website [Brait, 2015]. Therefore, the punishment of IBSA may be perceived as mild and non-severe. In addition, the felt-anonymity of cyberspace tends to ease one's concern with legal consequences and public stigmatisation [Akdeniz, 2002]. Perpetrators may perceive a low certainty in apprehension, especially when the cybercrime is transnational, as laws in the victim's residing country may not apply in the perpetrator's country.

Applying the Rational Choice Theory to our understanding of IBSA perpetrators' motivations, the perceived benefits of engaging in IBSA may outweigh the perceived deterrence, which then drives perpetrators into committing IBSA. The following section will thus look at the impact of IBSA followed by ways to mitigate it.

## 4.4 The Impact of Image-Based Sexual Abuse

As IBSA tend to be perpetrated online, the harm inflicted on victims can be magnified for various reasons [Citron & Franks, 2014]. First, the internet allows perpetrators to target their victims anonymously, which means that victims have to go through great pains to track perpetrators down and potentially seek redress [Franks, 2011]. Next, once information or sexual images of victims are posted on the World Wide Web, it is extremely difficult to remove all traces of them as they would be saved and redistributed on different websites and online platforms [Korenis & Billick, 2014]. Thirdly, victims may never "escape" their perpetrator due to "virtual captivity"—this means that even if the victim changes address, school, or workplace, he/she may still be harassed or "haunted" by the sexual images so long as the victim continues to maintain an online

presence [Franks, 2011]. Finally, sexually explicit photos or videos of victims have the potential to reach thousands, or even millions of netizens with just a click of the mouse. The audience can include friends, school mates, colleagues, and family members of the victim, as long as they have access to the Internet.

Taking these factors into consideration, this section discusses the impact that IBSA can have on victims—namely, reputational damage, threat to personal safety, and mental health issues.

### 4.4.1 *Reputational damage*

Firstly, one of the most observable impacts of IBSA on victims would be the threat to their reputation. Victims of IBSA often have their personal details posted online by the perpetrator alongside their sexually explicit materials. This is supported by the findings of Citron and Franks [2014], who reported that out of their study sample of 1,244 IBSA victims, more than 50% had their full names and links to their social media profiles posted with their nude photos or videos, while approximately 20% had their email addresses and phone numbers posted as well. The victim's personal information and sexual imagery may then pop up when potential employers attempt to do a background check on the victim, which is a common practice in the hiring process [Bloom, 2014]. For victims of relationship retribution IBSA, their sexual materials are often sent to their colleagues or superiors, who may then form ill impressions of the victim. Some victims of IBSA have faced very real consequences of professional and financial impact—with some victims losing their educational opportunities, being fired from their jobs, or losing clients and sales income [Bloom, 2014; Citron & Franks, 2014].

### 4.4.2 *Threat to personal safety*

With their private and identifiable information posted on the Internet and having no idea who has accessed these information, IBSA victims' physical and personal safety are compromised. Interviews with IBSA victims have revealed that victims do not feel safe to leave their homes after the victimisation as they are uncertain about their personal safety [Citron & Franks, 2014]. Victims also tend to face verbal harassment, physical harassment, and even stalking, which can affect their day-to-day

routine and functioning [Bloom, 2014]. A Singaporean victim of sexual voyeurism, Monica Baey, shared about the frustration and worry she was going through after the incident, to the extent of not feeling safe even in her own home [Ng, 2019].

### 4.4.3 *Mental health issues*

Victims of IBSA may also face a range of psychological effects from the experience. Victims who have had their sexually-explicit imagery shared without their consent reportedly suffer significantly worse mental health outcomes [Eaton *et al*., 2017]. In addition, victims have reported negative changes in their self-esteem and self-confidence after they were victimised—a change that can be attributed to the loss of control they experienced (i) during the actual act of IBSA; (ii) during police investigations and recovery; and (iii) as a result of the potential revictimisation [Bates, 2017]. Previous research in the area of self-esteem suggests that the failure to influence and control one's situation would lead to a sense of incompetency, inadequacy, and helplessness [Mruk, 2013]. These negative feelings could be experienced by IBSA victims due to the lack of autonomy in deciding who can view their naked body, loss of control over the propagation of the image, and uncertainty about who has gained access to the image [Bates, 2017; Crofts & Kirchengast, 2019].

In addition, a study by Ruvalcaba and Eaton [2020] found that victims of IBSA tend to have a lower psychological well-being, and may display more anxiety and depressive symptoms. Researchers have highlighted that IBSA may share similarities with the impact of sexual assault and sexual harassment—which means that high levels of stress, anxiety, and the possibility of post-traumatic stress disorder (PTSD) and depression observed among IBSA victims may mirror that of sexual assault survivors [Bloom, 2014; Gilboa-Schechtman & Foa, 2001; Woody & Beldin, 2012].

Some IBSA victims may develop suicidal thoughts as a result of the negative impact of victimisation on their mental health [Bloom, 2014]. According to a survey conducted by the Cyber Civil Rights Initiative between 2012 to 2013, approximately 51% of IBSA victims have contemplated suicide after the victimisation [Cyber Civil Rights Initiative, 2014]. Unfortunately, news of suicide among IBSA victims are not uncommon. For instance, a popular 28-year-old South Korean singer, Goo Hara, committed suicide in 2019, approximately a year after her ex-boyfriend

had threatened to release video footages of the both of them having sexual intercourse ["K-pop idol threatened with revenge porn," 2018; "K-pop star Goo Hara's suicide," 2019]. In another recent case in Korea, a bride-to-be committed suicide after being a victim of sexual voyeurism whereby she was secretly filmed in the changing room by a colleague ["Untouched yet ruined: Toll of South Korea spycam epidemic," 2020].

## 4.5 How can we Mitigate the Prevalence of IBSA?

### 4.5.1 *Identifying emerging trends in IBSA*

With vast improvements in technology, the modus operandi of IBSA perpetrators have evolved as well. An emerging trend of IBSA is sexualised photoshopping—the modification of a non-sexual photo to one of sexualised nature whereby the victim may be in various states of undress or depicted to be engaged in sexual activity [McGlynn & Rackley, 2017]. These are often done without the victims' knowledge through non-consensual downloading of an innocent photo the victim had posted online. With current technology, sexualised photoshopping can be done with just one click—by applications such as the Deep Nude app, which caused public furore when it was launched ["Shut down: DeepNude app," 2019]. Although taken down soon after its launch in 2019, the Deep Nude app allowed users to create a naked image of a fully-clothed woman by simply uploading the clothed photo onto the app [Cole, 2019].

Nevertheless, technological improvements have afforded perpetrators more than just photoshopping of images. A problematic trend of sexualised photoshopping includes the use of artificial intelligence to alter videos and create "deepfake porn" that look highly realistic to the untrained eye [McGlynn *et al*., 2019]. This allows perpetrators to depict victims in sex acts they did not engage in, which can cause greater harm and trauma to victims. Since it is extremely common for people to post images on social media platforms, anyone can fall prey to sexualised photoshopping. Therefore, strategies to mitigate IBSA need to take into account such emerging trends of IBSA.

### 4.5.2 *System-centric mitigation*

On the flipside, IBSA mitigation strategies could leverage on technological advancements. In recent years, social media platforms such as Facebook

has stepped up on preventing and removing non-consensual sexual materials from their online platform. In particular, Facebook included a button for users to report nude or offensive images, which are then flagged and subsequently removed [Facebook Community Standards, 2020]. In addition, with machine learning and artificial intelligence, Facebook detects and restricts postings of near nude and nude media that are shared non-consensually on their platforms—if the media is deemed to violate Facebook's community standards, it will be removed and the user's account may be disabled for sharing intimate content without consent [Davis, 2019]. Thereafter, Facebook uses photo-matching technology to prevent similar copies of the image from being shared by others on Facebook, Messenger, and Instagram which are all managed by them ["Facebook ramps up fight against 'revenge porn'," 2017]. In 2017, Facebook went one step ahead to combat IBSA with a proactive approach—via the implementation of the "Non-Consensual Intimate Image Pilot" in Australia and expanding it to the US, the UK, and Canada in 2018 [Waugh, 2018]. Through this pilot programme, users who wish to avoid being victims of IBSA can pre-emptively submit their own nude photographs to selected anti-revenge pornography advocacy groups in their country (which will then pass those photographs to Facebook), before they become targets/victims of IBSA on the online platform. Facebook then uses an algorithmic technology called perceptual image hashing [i.e., a reverse-image technology] to store the digital image hash[a] of the photo before deleting the original image from Facebook's server, in order to prevent and block the upload and spread of the image in the future [Romano, 2018].

Facebook's approach is an interesting and controversial way to tackle IBSA and the success of the pilot programme has encouraged more countries to be a part of the programme. Currently, countries such as Brazil, Mexico, Pakistan, and Taiwan are also on board the pilot programme ["Non-consensually shared intimate images pilot," 2020]. However, it may be challenging for potential victims to provide their own nude image to social media platforms, and not everyone may be receptive to the idea of sending their own nude images to a social media enterprise. Therefore, it is important for tech companies to partner law enforcement agencies in such prevention efforts. For instance, having designated

[a]Hash is a randomized code derived through cryptographic security method which turns information (e.g., images) into codes—a process that cannot be reversed.

officers within the Police Force or relevant authorities with the expertise of perceptual image hashing such that only the hashed versions of those files are sent to social media companies and online platforms to prevent uploads of the original nude files on these platforms. This means that potential victims only have to send their nude images to the authorities, and be assured that these images will not be sent to online platforms.

Next, as seen in the application of Rational Choice Theory in the earlier part of this chapter, deterrence against IBSA needs to be strengthened to further reduce or prevent acts of IBSA. In particular, the severity and certainty of punishment needs to be stringent, such that the deterrent effects of IBSA can outweigh the perceived benefits. This has been an issue that most victims of IBSA struggle with, as perpetrators are often seemingly let off with a light punishment, while some forms of IBSA such as sexualised photoshopping may not even be recognised as a crime in some countries [Gallagher, 2019]. In recent years, many countries have reformed their laws to adapt to the changing nature of technology-facilitated crimes. Nevertheless, these laws need to be strict because although no physical harm may be done to the victim, the psychological harm can be detrimental as aforementioned. Stringent deterrence can thus communicate a firm stance against IBSA to the public, which may help deter reoffending and prevent others from emulating acts of IBSA.

### 4.5.3 *Person-centric mitigation*

Public education can be another important means to mitigate the occurrence and impact of IBSA. Firstly, public awareness could be enhanced to highlight the prevalence of IBSA and emphasise the severity and certainty of the penalties meted out to perpetrators of IBSA, in order to deter others from offending.

Next, education efforts on IBSA such as campaigns and other innovative ways of engaging the public in the discussion of IBSA, can send an important message to the general public that IBSA is an important issue not to be dismissed. For instance, AWARE Singapore launched the "Taking Ctrl, Finding Alt" Contest in 2019 to encourage proposals from the public on ways to tackle IBSA. The proposed solutions are required to target prevention, support for victims, or better access to justice for victims, and the winning proposal will be given a S$6,000 grant to implement the solutions [AWARE, n.d.]. Such campaigns and competitions

may encourage the public to read up more on IBSA and see IBSA as a problem in society before coming up with suitable solutions to alleviate the issue of IBSA.

Also, through public outreach and education in sharing of experiences by past IBSA victims, other victims of IBSA can be educated on the potential legal recourse and resources in the community they may tap on to recover from the victimisation. Research by Ruvalcaba and Eaton [2020] on cyber dating abuse victimisation, which can include IBSA, found that approximately 73% of victims did not seek help after victimisation as most were embarrassed or afraid to do so. Some victims of IBSA may then turn to alcohol to cope with the stressful experience of perpetual rumination on the victimisation and possible escalation of the situation [Bates, 2017; Van Ouytsel *et al*., 2016]. Victims of IBSA tend to take a significant amount of time to gradually deal with the impacts of victimisation by seeking social support, religious help, being open to professional help such as counselling, and finding strength through victimisation to advocate for positive change in society [Bates, 2017]. These positive coping mechanisms were observed in the case of Holly Jacobs and Monica Baey—the former spoke out against IBSA by establishing the "End Revenge Porn" website [Jacobs, 2013], while the latter advocated against technology-facilitated sexual violence through interviews and by being part of the panellist for AWARE's campaign against sexual violence in 2019 [Gomez, 2019]. Therefore, the sharing of experiences by IBSA victims in support groups could help other victims seek positive coping mechanisms to deal with the impacts of victimisation.

Besides rendering help and support to the victims of crimes, it is also important to reach out to potential perpetrators who are at risk of committing IBSA. IBSA helplines can be set up for victims to seek emotional and informational support, and also to lend a listening ear for potential perpetrators to share their difficulties anonymously (e.g., a difficult break up and revenge plans, voyeuristic disorder, etc.). This may be a counselling helpline to steer potential perpetrators towards non-deviant methods of alleviating their existing issues. Such hotlines may be especially helpful in times of a pandemic, as evident from how the Revenge Porn Helpline in the UK saw a surge in traffic with more victims seeking help during the Coronavirus lockdown period, when stresses from break-ups and job loss could have aggravated IBSA perpetrators' pre-existing coercive control and abusiveness [Price, 2020].

## 4.6 Conclusion

In conclusion, Image-Based Sexual Abuse (IBSA) is an emerging form of cyber sexual deviant behaviour that has gained traction in recent years and may continue to evolve with new technological advances. With the popularisation of Virtual Reality (VR) technology, IBSA perpetrators could abuse such technological advances to cause more harm to their victims—however, more research will be needed to explore this possibility. This chapter has discussed the key typologies of IBSA—namely, relationship retribution, sextortion, sexual voyeurism, sexploitation, and the filming of sexual assault. Through the use of psychology, this chapter has highlighted the underlying motivation of IBSA perpetrators and the impact that IBSA has on its victims. By providing a psychological and humanistic view to the issue of IBSA, both system-centric and person-centric mitigation strategies were proposed to tackle IBSA. Although IBSA is an issue often examined from criminological and feministic angles, complimenting those perspectives with research on the psychology of perpetrators and victims of IBSA could help to mitigate this cyber-facilitated sexual crime in the long run.

## 4.7 Acknowledgement

The views expressed in this chapter are the author's alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 4.8 References

4 arrested for allegedly circulating obscene materials in SG Nasi Lemak Telegram chat group. (2019, October 15). *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/4-arrested-for-alleged-involvement-in-circulating-obscene-materials-in-sg

Akdeniz, Y. (2002). Anonymity, Democracy, and Cyberspace. *Social Research, 69*(1), pp. 223–237.

Alkhatib, S. I. (2018, May 4). Jail for man who pressed woman into choosing cash or her body to not spread nude photos. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/jail-for-man-who-pressed-woman-into-choosing-cash-or-her-body-to-not-spread

Alkhatib, S. (2019a, December 19). Nanyang polytechnic student admits taking upskirt photos, including at MRT stations. *The Straits Times.* https://www.

straitstimes.com/singapore/courts-crime/nanyang-polytechnic-student-admits-taking-upskirt-photos-including-at-mrt

Alkhatib, S. (2019b, November 28). Lecturer allegedly took upskirt videos of female victim in classroom. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/lecturer-allegedly-took-upskirt-videos-of-female-victim-in-classroom

Association of Women for Action and Research (AWARE). (2019). *AWARE sees overall rise in cases involving technology-facilitated sexual violence, launches contest to combat Image-Based sexual abuse.* https://www.aware.org.sg/2019/11/aware-sees-overall-rise-in-cases-involving-technology-facilitated-sexual-violence-launches-contest-to-combat-Image-Based-sexual-abuse/

AWARE. (n.d.). *Taking Ctrl, Finding Alt Contest*. https://www.aware.org.sg/tfsvcontest/

Bates, S. (2017). Revenge porn and mental health: A qualitative analysis of the mental health effects of revenge porn on female survivors. *Feminist Criminology, 12*(1), pp. 22–42. http://dx.doi.org/10.1177/1557085116654565

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime* (pp. 13–68). Palgrave Macmillan, London.

Bloom, S. (2014). No vengeance for 'revenge porn' victims: Unraveling why this latest female-centric, intimate-partner offense is still legal, and why we should criminalize it. *Fordham Urban Law Journal, 42*, pp. 233–289.

Boon, S. D., Alibhai, A. M., & Deveau, V. L. (2011). Reflections on the costs and benefits of exacting revenge in romantic relationships. *Canadian Journal of Behavioral Science, 43*, pp. 128–137.

Brait, E. (2015, December 3). Revenge pornography website operator sentenced to 25 months in prison. *The Guardian.* https://www.theguardian.com/us-news/2015/dec/03/revenge-pornography-website-operator-sentenced-isanyoneup-hunter-moore

Calvert, C., & Brown, J. (2000). Video voyeurism, privacy and the Internet: Exposing peeping toms in cyberspace. *Cardozo Arts and Entertainment Law Journal, 19*, pp. 469–566.

Chong, E. (2017, April 4). Man sent ex-girlfriend's nude photos to website. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/man-sent-ex-girlfriends-nude-photos-to-website

Citron, D. K., & Franks, M. A. (2014). Criminalizing revenge porn. *Wake Forest Law Review, 45*, pp. 101–140.

Cole, S. (2019, June 27). This horrifying app undresses a photo of any woman with a single click. *Vice.* https://www.vice.com/en_us/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman

Crofts, T., & Kirchengast, T. (2019). A ladder approach to criminalising revenge pornography. *The Journal of Criminal Law*, *83*(1), pp. 87–103. http://dx.doi.org/10.1177/0022018318814361

Cyber Civil Rights Initiative. (2014). *Revenge porn statistics.* https://www.cybercivilrights.org/wp-content/uploads/2014/12/RPStatistics.pdf

Cyber Civil Rights Initiative. (2020). *Definitions.* https://www.cybercivilrights.org/definitions/

Davis, A. (2019, March 15). Detecting non-consensual intimate images and supporting victims. *Facebook.* https://about.fb.com/news/2019/03/detecting-non-consensual-intimate-images/

Davies, S. (2020, May 6). Revenge porn soars in Europe's coronavirus lockdown as student fights back. *Reuters.* https://www.reuters.com/article/us-health-coronavirus-europe-porn-trfn/revenge-porn-soars-in-europes-coronavirus-lockdown-as-student-fights-back-idUSKBN22H2I6

Dir, A. L., & Cyders, M. A. (2015). Risks, risk factors, and outcomes associated with phone and internet sexting among university students in the United States. *Archives of Sexual Behavior, 44,* pp. 1675–1684. http://dx.doi.org/10.1007/s10508-014-0370-7

Durkin, K. F. (2001). Cyberporn and computer sex. *Encyclopedia of Criminology and Deviant Behaviour, 3*, pp. 62–66.

Durkin, K. F., Forsyth, C. J., Quinn, J. F. (2006). Pathological internet communities: A new direction for sexual deviance research in a post modern era. *Sociological Spectrum, 26*(6), pp. 595–606.

Dutton, M. A., & Goodman, L. A. (2005). Coercion in intimate partner violence: Toward a new conceptualization. *Sex Roles, 52*, pp. 743–756. http://dx.doi.org/10.1007/s11199-005-4196-6

Eaton, A. A., Jacobs, H., & Ruvalcaba, Y. (2017). *2017 Nationwide online study of nonconsensual porn victimization and perpetration: A summary report.* https://www.cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf

Facebook Community Standards. (2020). *9. Sexual exploitation of adults.* https://www.facebook.com/communitystandards/sexual_exploitation_adults

Facebook ramps up fight against 'revenge porn'. (2017, April 6). *The Straits Times.* https://www.straitstimes.com/tech/games-apps/facebook-ramps-up-fight-against-revenge-porn

Franks, M. A. (2011). Unwilling Avatars: Sexual Harassment in Cyberspace. *Columbia Journal of Gender and Law, 20*, pp. 226.

Gallagher, S. (2019, December 16). Government told revenge porn and deepfake laws 'not fit for purpose'. *Independent.* https://www.independent.co.uk/life-style/women/revenge-porn-deepfake-law-commission-review-a9248566.html

Gilboa-Schechtman, E., & Foa, E. B. (2001). Patterns of recovery from trauma: The use of intraindividual analysis. *Journal of Abnormal Psychology*, *110*, pp. 392–400. http://dx.doi.org/10.1037//0021-843X.110.3.392

Goh, Y. H. (2019, May 11). NUS student arrested after allegedly filming female student in bathroom. *The Straits Times.* https://www.straitstimes.com/singapore/nus-peeping-tom-case-another-female-student-allegedly-filmed-in-bathroom-male-student

Gomez, J. (2019, November 29). *A recap: Taking ctrl, finding alt 2019.* https://www.aware.org.sg/2019/11/a-recap-taking-ctrl-finding-alt-2019/

Grant, H. (2020, March 25). Urgent action needed as rise in porn site traffic raises abuse fears, says MPs. *The Guardian.* https://www.theguardian.com/global-development/2020/mar/25/urgent-action-needed-as-spike-in-porn-site-traffic-raises-abuse-fears-say-mps

Greenhouse, E. (2014, January 28). The downfall of the most hated man on the Internet. *The New Yorker.* https://www.newyorker.com/tech/annals-of-technology/the-downfall-of-the-most-hated-man-on-the-internet

Griffiths, M. (2000). Sex on the Internet. *In Children in the New Media Landscape: Games, Pornography, Perceptions. Edited by Feilitzen, C. V. and Carlsson, Ulla.* The UNESCO International Clearinghouse on Children and Violence on the Screen.

Henry, N., Powell, A., & Flynn, A. (2017). *Not Just 'Revenge Pornography': Australians' Experiences of Image-Based Abuse.* A Summary Report. Melbourne: RMIT University.

Holmes, S. T., & Holmes, R. M. (2008). *Sex crimes: Patterns and behaviour.* SAGE Publications.

Jacobs, H. (2013, November 27). Being a victim of revenge porn forced me to change my name. *Thought Catalog.* https://thoughtcatalog.com/dr-holly-jacobs/2013/11/being-a-victim-of-revenge-porn-forced-me-to-change-my-name/

K-pop idol threatened with revenge porn. (2018, October 8). *The Straits Times.* https://www.straitstimes.com/lifestyle/entertainment/k-pop-idol-threatened-with-revenge-porn

K-pop star Goo Hara's suicide sparks a reckoning on revenge porn, sexual assault in South Korea. (2019, November 29). *The Straits Times.* https://www.straitstimes.com/asia/east-asia/goo-haras-suicide-sparks-a-reckoning-on-revenge-porn-sexual-assault-in-south-korea

Korenis, P., & Billick, S. B. (2014). Forensic implications: Adolescent sexting and cyberbullying. *Psychiatric Quarterly, 85*(1), pp. 97–101.

Lam, L. (2019, November 26). Man who took upskirt videos as 'a challenge' sentenced to jail. *Channel News Asia.* https://www.channelnewsasia.com/news/singapore/man-who-took-upskirt-videos-as-a-challenge-sentenced-to-jail-12126096

Leaked sex tapes and child porn: A look into 13 illicit telegram chat groups. (2019, November 2). *Channel News Asia.* https://www.channelnewsasia.

com/news/singapore/telegram-chat-groups-nasi-lemak-porn-upskirt-sex-police-12054340

Loughran, T. A., Paternoster, R., Chalfin, A., & Wilson, T. (2016). Can rational choice be considered a general theory of crime? Evidence from individual-level panel data. *Criminology: An Interdisciplinary Journal, 54*(1), pp. 86–112. https://doi.org/10.1111/1745-9125.12097

Lum, S. (2018, March 28). Man who admitted to sexual offences faces new charges. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/man-who-admitted-to-sexual-offences-faces-new-charges

McAfee. (2013, February 4). Lovers beware: Scorned exes may share intimate data and images online. *McAfee.* Retrieved from http://www.mcafee.com/au/about/news/2013/q1/20130204-01.aspx

McCoy, T. (2014, September 2). 4chan: The 'shock post' site that hosted the private Jennifer Lawrence photos. *The Washington Post.* Retrieved from https://www.washingtonpost.com/news/morning-mix/wp/2014/09/02/the-shadowy-world-of-4chan-the-shock-postsite-that-hosted-the-private-jennifer-lawrence-photos/

McGlynn, C., & Rackley, E. (2017). Image-Based Sexual Abuse. *Oxford Journal of Legal Studies, 37*(3), pp. 534–561. https://doi.org/10.1093/ojls/gqw033

McGlynn, C., Rackley, E., & Houghton, R. (2017). Beyond 'revenge porn': The continuum of Image-Based sexual abuse. *Feminist Legal Studies*, *25*(1), pp. 25–46.

McGlynn, C., Rackley, E., Johnson, K., Henry, N., Flynn, A., Powell, A., Gavey, N., & Scott, A. (2019). *Shattering lives and myths: A report on Image-Based sexual abuse.* Durham University; University of Kent Report.

Miller, M. E. (2013, May 2). Miami revenge porn victim Holly Jacobs demands politicians "take issue seriously." *Miami New Times.* https://www.miamine wtimes.com/news/miami-revenge-porn-victim-holly-jacobs-demands-politicians-take-issue-seriously-6520948

Ministry of Law. (2019). *Commencement of Amendments to the Penal Code and Other Legislation on 1 January 2020 [Press Release].* https://www.mlaw.gov.sg/news/press-releases/commencement-of-amendments-to-the-penal-code-and-other-legislation-on-1-january-2020

Morris, A. (2012, November 13). *Hunter Moore: The most hated man on the Internet.* https://www.rollingstone.com/culture/culture-news/hunter-moore-the-most-hated-man-on-the-internet-184668/

Mruk, C. J. (2013). *Self-esteem and positive psychology: Research, theory, and practice*. Springer Publishing Company.

Ng, H. (2019, May 2). NUS peeping tom case: Monica Baey urges online bullying against Nicholas Lim to stop. *Asia One.* https://www.asiaone.com/singapore/nus-peeping-tom-case-monica-baey-urges-online-bullying-against-nicholas-lim-stop

Nigam, S. (2018, April 25). Revenge porn laws across the world. *The Centre for Internet & Society.* https://cis-india.org/internet-governance/blog/revenge-porn-laws-across-the-world#_Toc511943104

Nolan, T., & Maguire, M. (2016). Sex Offenders and their Treatment. In *Sex, Sexuality, Law, and (In) justice* (pp. 402–435). Routledge.

Non-consensually shared intimate images pilot. (2020). *Facebook.* https://www.facebook.com/safety/notwithoutmyconsent/pilot/partners

O'Hara, A. (2019). *Retribution-style adult image-based sexual abuse: crime scripting, CRAVED and situational crime prevention* (Doctoral dissertation, The University of Waikato).

Powell, A. & Henry, N. (2015). *Digital Harassment and Abuse of Adult Australians. A Summary Report.* Melbourne: RMIT University.

Powell, A., & Henry, N. (2017). *Sexual Violence in a Digital Age*. Basingstoke: Palgrave Macmillan.

Powell, A., Henry, N., & Flynn, A. (2018). Image-based sexual abuse. *Handbook of critical criminology*, pp. 305–315.

Powell, A., Henry, N., Flynn, A., & Scott, A. J. (2019). Image-based sexual abuse: The extent, nature, and predictors of perpetration in a community sample of Australian residents. *Computers in Human Behavior, 92,* pp. 393–402.

Price, H. (2020, April 25). Coronavirus: 'Revenge porn' surge hits helpline. *BBC News.* https://www.bbc.com/news/stories-52413994

Quinn, J. F., & Forsyth, C. J. (2005). Describing sexual behaviour in the era of the internet: A typology for empirical research. *Deviant Behavior, 26*, pp. 191–207. https://doi.org/10.1080/01639620590888285

Romano, A. (2018, May 24). Facebook's plan to stop revenge porn may be even creepier than revenge porn. *Vox.* https://www.vox.com/2018/5/23/17382024/facebook-revenge-porn-prevention

Ruvalcaba, Y., & Eaton, A. A. (2020). Nonconsensual pornography among US adults: A sexual scripts framework on victimization, perpetration, and health correlates for women and men. *Psychology of violence*, *10*(1), pp. 68. http://dx.doi.org/10.1037/vio0000233

Salter, M., & Crofts, T. (2015). Responding to revenge porn: Challenges to online legal impunity. In L. Comella & S. Tarrant (Eds.), *New views on pornography: Sexuality, politics, and the law* (pp. 233–256). Praegar.

Shut down: DeepNude app that can virtually 'undress' women. (2019, June 30). *The Straits Times.* https://www.straitstimes.com/world/shut-down-deepnude-app-that-can-virtually-undress-women

Sim, D. (2019, April 19). Singapore student Monica Baey wants firm action from NUS after man who filmed her in hostel shower goes 'scot-free'. *South China Morning Post.* https://www.scmp.com/news/asia/southeast-asia/article/3006950/singapore-student-monica-baey-lashes-out-nus-after-man-who

Singapore Legal Advice. (2020, February 7). *Revenge Porn: What if your nudes are leaked in Singapore?* https://singaporelegaladvice.com/law-articles/revenge-porn-nudes-leaked-singapore/

Sirianni, J. M., & Vishwanath, A. (2016). Bad romance: Exploring the factors that influence revenge porn sharing amongst romantic partners. *Online Journal of Communication and Media Technologies*, *6*(4), p. 42.

Skinner, B. F. (1938). *The behaviour of organisms.* New York: Appleton-Century

Su, R., Porter, T., & Mark, M. (2019, October 30). Here's a map showing which US states have passed laws against revenge porn—and those where it's still legal. *Business Insider US.* https://www.businessinsider.sg/map-states-where-revenge-porn-banned-2019-10?r=US&IR=T

Revenge porn: What is the law in the UK? (2019, June 26). *The Week.* https://www.theweek.co.uk/101962/revenge-porn-what-is-the-law-in-the-uk

Thompson, M. P., & Morrison, D. J. (2013). Prospective predictors of technology-based sexual coercion by college males. *Psychology of Violence, 3*, pp. 233–246. http://dx.doi.org/10.1037/a0030904

Uhl, C. A., Rhyner, K. J., Terrance, C. A., & Lugo, N. R. (2018). An examination of nonconsensual pornography websites. *Feminism & Psychology*, *28*(1), pp. 50–68. http://dx.doi.org/10.1177/0959353517720225

Untouched yet ruined: Toll of South Korea spycam epidemic. (2020, January 13). *The Straits Times.* https://www.straitstimes.com/asia/east-asia/untouched-yet-ruined-toll-of-south-korea-spycam-epidemic

Van Ouytsel, J., Ponnet, K., Walrave, M., & Temple, J. R. (2016). Adolescent cyber dating abuse victimization and its associations with substance use, and sexual behaviors. *Public Health, 135*, pp. 147–151. http://dx.doi.org/10.1016/j.puhe.2016.02.011

Wahlquist, C. (2019, May 6). First person charged under Western Australia's new revenge porn laws. *The Guardian.* https://www.theguardian.com/australia-news/2019/may/06/first-person-charged-under-western-australias-new-revenge-porn-laws

Waugh, R. (2018, May 23). Facebook wants you to send in your nudes so it can 'protect' you. *Metro News.* https://metro.co.uk/2018/05/23/facebook-wants-send-nudes-can-protect-7570275/

Woody, J. D., & Beldin, K. L. (2012). The mental health focus in rape crisis services: Tensions and recommendations. *Violence and Victims, 27*(1), pp. 95–108.

# Section C

# Hate Crimes Online

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0005

# Chapter 5

# "Is Technology Making You Prejudiced?": How Technology is Enabling Hate IRL

Nur Aisyah Abdul Rahman

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Aisyah_ABDUL_RAHMAN_from.TP@mha.gov.sg*

## 5.1 Introduction

In 2017, a photo of a Muslim woman in the initial aftermath of the Westminster Bridge attack[a] in London went viral, particularly among right-wing[b] circles. The photograph depicted a Muslim woman, seemingly walking past and ignoring an injured victim, while the victim was lying on the ground and being helped by others as emergency services were

[a]On 22 March 2017, Briton Khalid Masood rammed a car into pedestrians along the Westminster Bridge and stabbed a police officer near the Houses of Parliament. The attack was classified as an Islamist-related act of terror. Five people were killed while 49 others were injured (Allen & Henderson, 2017).

[b]Right-wing extremism does not have a specific definition, but its main themes tend to revolve around ethnocentrism and an intolerance to certain groups (Mudde, 2000). It has evolved over the years—beginning with fascism, Nazism, and White supremacy, it has evolved to become more anti-immigrant, xenophobic, and Islamophobic in recent years (Baysinger, 2006; Hafez, 2014).

presumably on their way. The photograph was subsequently lifted out of context, paired with inflammatory captions to perpetuate hate. For example, one account tweeted, "Muslim woman pays no mind to the terror attack, casually walks by a dying man while checking phone #PrayForLondon #Westminster #BanIslam" [Mortimer, 2017, para 3]. The picture eventually became a right-wing meme meant to perpetuate malicious Islamophobic sentiments online.

As social beings, the internet and various social media platforms embody a virtual space where people can interact, very different from physical and face-to-face interactions. This is in part due to the advent of technology, which has allowed people to connect with others in unprecedented ways. Now, communication has transcended geographical constraints, connecting people based on their interests. Due to its interactive nature, certain technological advances (e.g., algorithms) have been exploited to manipulate human behaviours. As this chapter will discuss in detail later, the online environment can be a place that encourages cyberhate, while acknowledging that humans have the agency to choose how they behave in certain online networks.

The Anti-Defamation League (ADL)[c] defines cyberhate as:

> "any use of electronic communications technology to spread anti-Semitic, racist, bigoted, extremist or terrorist messages or information. These electronic communications technologies include the Internet (i.e., Web-sites, social networking sites, 'Web 2.0' user-generated content, dating sites, blogs, on-line games, instant messages, and E-mail) as well as other computer- and cell phone-based information technologies (such as text messages and mobile phones)." [ADL, 2010, p. 1]

While the internet and various online platforms are designed to be interactive tools that help fulfil the human need to socialise, this chapter discusses how cyber perpetrators have utilised these interactive tools to reinforce prejudiced beliefs through spreading cyberhate. We will cover four characteristics of the online environment that cyber perpetrators have exploited to spread hate. This chapter will also suggest a possible variety of systemic and human interventions to combat cyberhate.

---

[c] The ADL is an international non-governmental organization based in the United States of America, which seeks to fight against anti-Semitism and bigotry (ADL, n.d.).

## 5.2 The Online Environment Encourages People to be Disinhibited and Thus More Likely to Express Cyberhate

Online environments encourage psychological distance and reduced accountability, emboldening individuals who do things that they might not necessarily do in real life (IRL) [Suler, 2004]. Cloaked under a sense of anonymity and invisibility, individuals might be more likely to express hateful sentiments, encouraged by the fact that they can appear as whomever they want to be online.

Suler [2004] purported that this perceived anonymity facilitates a reduction in perceived accountability of one's actions online. Referred to as the disinhibition effect, it occurs because people are able to separate their online behaviours from their offline self. Lowry *et al.* [2016] further elaborate on the role of anonymity online by arguing that the perceived anonymity online is facilitated by the following factors: the lack of identification (i.e., the perception that personal identities will not be revealed), diffused responsibility (i.e., the belief that one will not be held accountable for one's online actions), lack of proximity (i.e., the perception that no one else is physically close to observe one's online deviant online actions), lack of knowledge of others (i.e., the belief that others online are not able to recognise them as the same person), and confidence in the system to function (i.e., having confidence in the fact that the online platforms will not malfunction and reveal one's real identity).

In one study, Wachs and Wright [2018] found a positive correlation in German students who reported higher levels of online disinhibition and their frequency of online hate perpetration (e.g., posting hateful comments attacking minority group members).

Disinhibition, however, is not necessarily a bad thing. Online disinhibition can empower people as well. For example, in 2016, Sarah Camariah shared her experience of racial discrimination in a Facebook post, describing how she was racially stereotyped during her job interview at a halal-certified Singaporean bakery chain, PrimaDeli. Seeing that she was ethnically Malay, the interviewer told her "You know ah, Malays ah they over promise, promise I can do this I can do that, in the end, cannot make it, after 2 days disappear" [Sarah, 2016, para 7]. Sarah explained that she shared her story in order to raise awareness about racial discrimination when applying for jobs, even at a company that caters to

the Muslim and Malay community. The post went viral, and PrimaDeli eventually dismissed the staff in question [Ho & Aw, 2016]. For Camariah, social media had helped her to find a voice to spread awareness on racially discriminatory practices in Singapore.

Conversely, in cases such as the Westminster Bridge photo, disinhibition prompted a malicious consequence. Particularly with the ability to be anonymous or invisible, for some, online platforms evolve to embody a space in which they are able to express their worst self. In one example in Singapore, a thread was started on a *HardwareZone*[d] forum, discussing a photo that a male student had posted on social media of him kissing his male partner [Daud, 2019]. Some of the participants of the thread responded with homophobic and disconcerting comments that called for violence. One participant going by the username *UptheToon*, commented, "These chao [smelly] faggots getting more daring. Put them in gas Chambers and clean them." Some of the other participants went even further by identifying the student and lodging a complaint with the student's school [Daud, 2019]. Recently, Abdul Halim Abdul Karim, an Islamic religious teacher in Singapore, claimed in a Facebook post that the COVID-19 outbreak was retribution for China's oppressive treatment of Chinese Muslim Uighurs [Kurohi, 2020]. Although Abdul Halim has since publicly apologised and clarified that he did not intend the post to be a racial attack [Kurohi, 2020], for some netizens, his claims were an attack on people of Chinese ethnicity. It is important to note that Abdul Halim used online spaces to unleash loaded prejudiced opinions. Such homophobic and racial attacks are examples of how people utilise online spaces the wrong way, expressing impolite opinions liberally that they would never have dared to do in real life.

## 5.3 Cyber Perpetrators Utilise Algorithms to Reinforce Prejudicial Beliefs

Ever wondered how YouTube or Netflix recommend videos that you might like? Or how Facebook ads seem to know that you are interested in buying jewellery or going on a beach weekend getaway? A lot of what appears on the screen of one's device is due to machine learning

[d]HardwareZone is an IT-oriented website, with the forum being commonly used to discuss a variety of issues in Singapore.

algorithms. One of the ways which how machine learning algorithms work is by utilising user-data. Machine learning algorithms gather the data that Internet users make available on the Internet (e.g., visited websites, liked posts, search terms, the amount of time spent on certain sites) and compare that to other people's data that are similar to the user's [Fakhfakh *et al*., 2017]. They then use this to predict what that user might like based on what other people with similar behaviours have liked [Fakhfakh *et al*., 2017]. Essentially, this algorithm process is similar to asking a friend who shares the same interests as you for recommendations, except that algorithms extrapolate recommendations using data interpolated from thousands of similar individuals instead. Algorithms also provide recommendations by identifying similar characteristics across different products that a user have shown interest in [Fakhfakh *et al*., 2017]. For example, if you watched *Mission Impossible* on Netflix, the system is likely to suggest one of the sequels or another movie in the spy or action genre or another movie starring Tom Cruise. Some of the user data is also shared with other parties [Zang *et al*., 2015], resulting in user-personalised advertisements on a different website or mobile app. One study found that certain applications shared identifying information (e.g., email addresses), location data (e.g., geo-coordinates), and behavioural information (e.g., search terms) with other organisations [Zang *et al*., 2015]. This information can be used to infer personal details (e.g., one's religion based on places of worship visited) about the user and target specific messages. In one example, an advertising company used phone location data to target young women who were near reproductive health clinics with anti-abortion advertisements on their phones [Woodward & Bray, 2017].

Machine learning algorithms are inherently biased because they work based on the pattern of examples that are provided by its human programmers, who inform the algorithm on what they should pay attention to and who to target [Fakhfakh *et al*., 2017]. In 2015, Google came under scrutiny when its machine learning algorithm classified photos of African Americans as *gorillas* in Google Photos [Simonite, 2018]. Label searches for 'black person', 'black woman', or 'black man' led to the retrieval of black and white photos of people categorised by the right gender but not the right race. Reminiscent of racist tropes that dehumanises people of African descent by likening them to primates [Staples, 2018], this algorithm had inadvertently reinforced racial prejudices. Google later apologised and explained that the mistake was a result of limitations in their algorithm programming [Hern, 2018; Simonite, 2018]. However,

critics claimed that such inherent biases are due to lack of diversity in tech companies that build these algorithms [Pulliam-Moore, 2015]; the algorithm is only as diverse as how the programmers teach it to be.

While the gorilla debacle was an unwitting mistake, algorithms can be used deliberately to identify users who are susceptible to specific messages, which was the case of the Facebook-Cambridge Analytica data scandal in 2018. Facebook had left the data of users (and their friends') vulnerable to third-party applications [Meredith, 2018; Richterich 2018]. Some of these applications gathered personal information not only from direct users (i.e., those who used the app), but also those who were in their social network (i.e., all the users' friends on Facebook). Through a quiz application, Cambridge Analytica gained access to data from 87 million users [Chang, 2018; Richterich 2018]. With identities and data on users' social media network (i.e., list of friends and circle of influence) as well as their online interactions (e.g., what do they like), Cambridge Analytica was able to build aggregated 'psychographic profiles' of these users [Meredith, 2018]. This information was then allegedly used to identify specific users to direct pro-Brexit and pro-Trump messages to, in order to sway voters during elections [Meredith, 2018]. Considering the voting results, Cambridge Analytica's ethically questionable practices had come under public scrutiny and shone a light on user data security. Algorithms that steal and utilise user data can reinforce people's biases. These biases can take on a fanatical form if extended to religious or political beliefs that people tend to feel very strongly about.

As social media users, people are susceptible to the availability heuristic or bias. The availability heuristic is a tendency whereby people rely on whatever salient information that comes to their minds about a specific topic [APA, n.d.]. The prevalent use of machine learning algorithms, coupled with humans' susceptibility to the availability heuristic, increases any vulnerability to prejudiced messages and cyberhate. As in the case of the Westminster attack, the inflammatory meme was started by a Twitter account which had right-wing extremist, anti-Semitic, and Islamophobic content [Hern, 2017]. With engagements with other right-wing accounts and adopting the use of popular right-wing hashtags, the account and its tweets were both primed to appear on other right-wing supporters' Twitter feeds and social media networks. Seeing the same content and perspective appear repeatedly on their social media feed is bound to prime people into thinking that that is a popular opinion or that

a particular incident is likely true. Although the photographer of the Westminster attack photo had initially used it to illustrate the distress in the aftermath of the attack**,** the prevalence of posts that claimed otherwise had effectively changed the narrative and encouraged further Islamophobia. Moreover, it was also revealed that one of the most influential accounts responsible for the Westminster photo was, in fact, a troll account that was part of the Russian disinformation campaign, to stir tensions in the United States and the United Kingdom [Dixon, 2017; Hern, 2017]. Hence, not only can algorithms be used to connect similar people and content; they can also be used to proliferate prejudice.

## 5.4 Online Communities are Echo-Chambers where Biased Beliefs are Reinforced

Online communities can be echo-chambers that also reinforce divisions in society. More often than not, people fall prey to their own confirmation bias and selectively adhere to information that is congruent to their beliefs [Bright, 2017; Del Vicario *et al*., 2016]. Similar to offline group dynamics, the constant reinforcement of similar sentiments on social media enables a collective identity and solidarity [Hanzelka & Schmidt, 2017]. Echo-chambers or extreme stances can thus facilitate the establishment of a group identity. Such group identities are problematic when they encourage prejudice. People with more extreme viewpoints and have high confidence in their stance are also more likely to selectively expose themselves to confirmatory information, and remain unmoved by alternative viewpoints [Bright, 2017]. Contrary to the belief that dissent would provide an alternative perspective, research on group dynamics shows that group polarisation towards an extreme opinion is likely to happen especially after a group discussion on that topic [Sunstein, 2007]. In one study, Yardi and Boyd [2010] found that discussion of highly politicised topics on Twitter strengthened group identity; discussion between like-minded people reinforced ingroup (i.e., people who are perceived to be part of the same social group) identity and beliefs; while discussions with different-minded people reinforced an 'Us versus Them' perspective. It is possible that individuals feel pressured to conform to the perceived group norm (i.e., the popular viewpoint, even if it is extreme) in order to reinforce their group identity and feel accepted by other group

members [Bright, 2017; Sunstein, 2007]. Another reason why group polarisation occurs is that people refer to others as information sources to corroborate their existing knowledge [Sunstein, 2007]. Additional information from others might push someone who was ambivalent towards a particular stance.

Considering that the algorithms are going to consistently recommend similar content or other people who like the same content (e.g., "People You May Know" suggestions on Facebook), users could end up in an echo-chamber of like-minded people. In some cases, people actively seek echo-chambers by seeking out people who they know share similar views through social media groups and websites that cater to specific causes. With tightening regulations against hate speech on mainstream social media, right-wing advocates have sought refuge in alternative forums (e.g., 8Chan[e]) and social media platforms (e.g., Gab[f]) as 'safe spaces' to express themselves. Social media is a pervasive medium that violent extremists use to spread their rhetoric and identify people to recruit [Stalinsky *et al*., 2016]. On these platforms, the group is often portrayed to be under attack by the target group [Bartlett & Birdwell, 2013]. For example, one of most the common rhetoric of current right-wing violent extremists is that the Muslims are out to eradicate the Western world, with Islamic State of Iraq and Syria (ISIS) at the helm and enabled by all those who are pro-multiculturalism [Bartlett & Birdwell, 2013]. Much of Islamic violent extremist propaganda contains the same fearmongering and demonising of the Western world as the oppressors, Muslims as victims, and pro-multicultural moderates as enablers [ADL, 2016]. The confidence and steadfastness with which echo-chambers hold onto their beliefs emphasise the divisions in society.

With the major mainstream tech companies such as Facebook and Twitter tightening regulations against extremism online and shutting down extremist accounts, these violent extremist groups have migrated to more secure platforms. In some cases, these echo-chambers come in the form of exclusive online chat groups on encrypted messaging platforms

[e] 8Chan is a message board forum website that is popular with right-wing violent extremists due to features that enable a sense of community while still keeping users anonymous. The website has been linked to multiple attacks (e.g., Christchurch mosque shootings, 2019), where perpetrators have posted their manifestos prior to their attacks (Roose, 2019).

[f] Gab is an alternative social media platform with a largely right-wing user base (Hall, 2019). It is characterised as a pro-free speech and uncensored platform (Hall, 2019).

such as WhatsApp and Telegram. Telegram, in particular, has become popular among violent extremists. ISIS and their supporters are known to disseminate propaganda and even terror attack plans and manuals on Telegram chats [Stalinsky *et al*., 2016]. Propaganda and some Twitter-ready posts are made available for members to copy and paste onto their own newsfeed [Tan, 2017]. Some of the functions of Telegram include temporary secret chats with self-destruct timers (i.e., think of the online equivalent of *Mission Impossible* mission notification that Tom Cruise's character gets but without the small self-destruct explosions) and encryption keys that allow users to assess if their communication is still secure. Such messaging platforms also have the capability to conceal members' identities. Unlike social media platforms such as Twitter and Facebook, where information on the network is publicly available, these chats make the hate networks more challenging to trace. With such capabilities, Telegram has also become increasingly popular among right-wing violent extremists [Hayden, 2019; Owen, 2019]—ironically, this is another commonality that this group of people share with their biggest ideological adversaries, ISIS.

In some cases, violent extremists create their own platforms in order to minimise the risks of getting removed by the platform management or, to further safeguard their data. For example, ISIS has created various applications to keep supporters updated with their latest news, such as *The Dawn of Glad Tidings* news app [Awan, 2017]. Right-wing news applications and websites, such as *Ritam* in India and *InfoWars* in the United States, further skew consumers' perception of the world by portraying a highly prejudiced perspective of global current affairs. Frequent and biased updates of violent extremist groups can also portray a much more powerful image of the group than in reality [Awan, 2017]. Not only would such psychological warfare tactics work to garner more support, it also reinforces their adversaries' claims that they are a relevant threat to defend against.

The above examples highlight how information on the Internet can shape people's views and attitudes. The presence of like-minded others can encourage individuals to perpetuate cyberhate. Research has found that perceived anonymity in a group and social diffusion can embolden individuals with a mob-mentality [Lowry *et al*., 2016; Seaman, 2008]. In other words, individuals might feel that they can get away with their deviant online behaviours and not be personally identified because they think that there are many others like them.

## 5.5 The Cyber World Provides a Platform to Commit Offline Hate

As much as the cyber sphere may offer anonymity and distance, not all things remain online. Cyberhate can have very real offline consequences. The omnipresence of the Internet, and by extension, the cyber audience, can make victims of cyber hate feel constantly targeted. Indeed, cyberhate can have detrimental consequences on one's mental and physical well-being [Awan & Zempi, 2015]. In the Westminster photo incident, the woman reported distress and fear at her pictures being circulated online [TellMAMA, 2017]. With the increasing negative attention, she had felt compelled into clearing up any misconceptions about the photo. Such fears are not unfounded as *doxxing* (i.e., the publishing of one's personal details with malicious intent) have led to distressing offline consequences [Tam, 2018], such as in the HardwareZone case.

Cyberhate can also translate to offline hate and vice versa. Research on hate crimes after terror attacks show that offline hate crimes and cyberhate increase after an attack [Evolvi, 2018]. Crises such as terror attacks and health pandemics are 'trigger' events that hate perpetrators can take advantage off to encourage dissent [Awan & Zempi, 2015; Hanzelka & Schmidt, 2017]. Thanks to the internet and globalisation, the impact of crises is not geographically constrained. In the case of the Charlie Hebdo attacks in Paris, there was an increase in reports of anti-Muslim hate crimes all over Europe [Awan & Zempi, 2015]; after the Paris attacks in 2015, #KillAllMuslims trended in the United Kingdom on social media, with some posts calling for the genocide of Muslims worldwide [Evolvi, 2018]. More recently, as the COVID-19 outbreak became a global pandemic, one study reported a 900% increase in hate speech directed at Asians and China on Twitter, with common hashtags such as *Kungflu*, *Chinesevirus, Wuhanvirus, Chinaliedpeopledied* [L1ght, 2020]. Globally, there have been increasing reports of Asians being the target of hate crimes and blamed for the outbreak [Yong, 2020].

Prolonged exposure to cyberhate desensitises individuals, making people perceive it to be normative and acceptable [Wachs & Wright, 2018]. Research that looks into the pathways of hate-based violence, such as ADL's Pyramid of Hate, purport that less severe expressions of hate (e.g., stereotyping, bullying, microaggressions) precede more severe expressions (e.g., hate crimes, terrorism, genocide). These less severe expressions of hate are normalised and accepted, thus desensitising

individuals and gradually enabling them to partake in increasingly severe behaviours [ADL, 2018]. Cyberhate is a form of aggression with very real psychological consequences, such that when these attitudes and behaviours are normalised, it can embolden certain individuals to act and commit more harmful forms of offline hate. Being in an echo-chamber can also encourage certain individuals to offline action. As depicted in the HardwareZone forum example, not only do like-minded others tend to validate each other's feelings, there may also be those who encourage others to commit violence. Lone wolf actors such as right-wing violent extremist Anders Behring Breivik [i.e., Oslo and Utoya attacks; Ravndal, 2013] and Islamic violent extremist Omar Mateen [i.e., the Pulse Nightclub attack; Antinori, 2017], were radicalised by prejudiced material and like-minded people online.

The cyber realm provides like-minded extremists the ability to both connect online and offline. Research has found that mobilisation of radicalised individualised is empirically more significant on Facebook than email [Hanzelka & Schmidt, 2017]. Social media platforms are also more effective at identifying potential supporters and spreading radical rhetoric [Hanzelka & Schmidt, 2017]. Right-wing extremist groups such as Britain's First and Pegida have been known to use social media to spread their rhetoric and plan protests [Evolvi, 2018].

Social norms, even those that are online, are becoming more egalitarian. People who are perceived to violate these online rules are punished by the majority, and at times, by their corresponding offline community. In one such case in Singapore, a video of a young Chinese woman in a ride-hailing app vehicle claiming that she was being held hostage and racially discriminated against by a Malay driver went viral in 2019. The video was interpreted by most netizens in favour of the driver, while the woman was deemed ignorant and entitled. The video garnered plenty of attention and even spurred the development of memes and parodies by various individuals and organisations, including the Singapore Civil Defence Force. The intense amount of negative attention and the eventual identification of the woman in the video led her to delete her social media accounts. There were also netizens who exposed (*doxxed*) her personal details online by revealing her full name and place of work [Lim, 2019; Shanmuganathan, 2019]. Citizen journalism media outlet, *The Temasek Review*, claimed that reporters had allegedly tried to find her at her workplace, only to discover that she fled overseas to escape the attention [The Temasek Review, 2019]. When netizens doxxed her

identity, leading to reporters appearing at her workplace, this cyber aggression then transpired to offline harassment.

## 5.6 Interventions

Thus far, this chapter has identified technological and human characteristics which enable cyberhate. Since the internet and social media are highly interactive technologies, interventions would need to address both the system and the user.

### 5.6.1 *Leverage on technology and technological companies to reduce user susceptibility to cyberhate*

From cases such as the Facebook-Cambridge Analytica scandal, social media and technology companies need to do more to safeguard user data. On the back end, they should ensure that they do their part in minimising data breaches. On the front end, tech companies need to be more transparent in communicating what user data is used for. For example, Facebook has been utilising pop-up notifications that inform users of data sharing with third-party organisations, including which data is specifically used. To improve transparency even further, tech companies can also include an opt-out feature (i.e., users can choose to be excluded from certain default features) for machine-learning algorithm suggestions or user data sharing with third parties. Such customisable features will allow consumers to have better control and security over their data. These systemic changes are important measures in reducing users' vulnerabilities.

Another way to leverage on technology as an intervention is to use algorithms to flag up hate material for removal. A group of researchers from McGill University in Canada have found that using hate speech filters for separate target groups (e.g., women, Asians) on *Reddit* were more accurate than simply identifying the use of certain words [Griffiths, 2019]. Teaching an algorithm to recognise hate is difficult due to language inconsistencies (e.g., spelling differences, grammar mistakes) and contextual factors that can imply different sentiments. While such research is still in its nascence, the progress that has been made is promising.

### 5.6.2 *Implementing laws and policies that target biased algorithms*

Algorithms are inherently biased and require human oversight to prevent offline discrimination. In targeted marketing, advertisers are able to target specific demographics to receive their advertisements. For example, one advertiser ProPublica, had reportedly requested that Facebook not show the organisation's housing rental advertisements to users who were "African Americans, mothers of high school kids, people interested in wheelchair ramps, Jews, expats from Argentina and Spanish speakers" [Angwin *et al*., 2017, para 3]. This was in spite of Facebook's allegations of stricter enforcement against discriminatory advertisement practices [Angwin *et al*., 2017]. One study found that even when advertisers tried to reach a more inclusive audience on Facebook, the algorithm sent out the advertisement to specific users whom the system identified as relevant [Ali *et al*., 2019].

As such, it is important that there are laws and policies in place that regulate the implementation of biased algorithms as well as call for greater accountability of organisations who use them. In Singapore, laws, such as the Protection from Online Falsehoods and Manipulation Act (POFMA) and the Sedition Act, help to deter and mitigate hate speech and fake news. While these laws might apply to users who create and/or spread cyberhate, there are not many laws or policies that regulate biased algorithms. In the United States, the Algorithmic Accountability Act is a proposed[g] bill that aims to increase organisations' accountability over the implementation of unlawful and unethical algorithms [Robertson, 2019]. It would mandate organisations to assess if their algorithms are biased or pose privacy or security risks to consumers.

### 5.6.3 *Reducing impulsive acts and creating a safe online space*

Brown [2018] suggested that instantaneousness is one of the distinguishing factors of online hate from its offline counterpart. In real life, individuals need to be well-prepared if they were to conceal their identity

[g]Accurate at the time of writing this chapter.

(e.g., mask, target that does not know them, etc.) and get away with a hate crime. It is however, much easier to conceal one's identity in the online sphere and get away with cyberhate (e.g., use temporary usernames, anonymous forums). Hence, cyberhate is more likely to be an act of impulse than offline hate. As mentioned earlier, there are already some algorithms built to recognise hate speech. These algorithms can be used on social media platforms to alert users that their post may contain hate speech before they publish their posts. This then leads to an extra step that users have to undergo before posting. For example, there can be pop-ups with a message that nudges at a positive social norm, "Are you sure you are ready to post? Does your post comply with the guidelines?," which users need to click on. The idea here is to give users a pause and not impulsively post potentially hateful content. At the same time, it can leverage on group membership and norms to reduce deviant behaviours online.

### 5.6.4 *Encourage users to be critical and aware*

There needs to be more awareness and transparency on how people can be targeted online. While more social media companies do their part in informing users on their information security, users also need to do their part in reading terms and conditions before they give apps and websites access to their personal information. Campaigns can be done to educate people on such cyber threats and what cyber hygiene habits can help. For example, the Singapore National Library Board conducts workshops and provides free resources on combating fake news through their S.U.R.E. outreach programme. With such awareness programmes, individuals will hopefully be more cognisant and critical of the information that they see online. This can potentially reduce their vulnerability to being targeted by algorithms and relying on availability heuristics.

Considering that increasingly more youths are spending more time online and the inclusion of coding in the Singaporean academic curriculum, it is essential to complement this with good cyber hygiene practices in the curriculum as well. The Massachusetts Institute of Technology (MIT) developed an open-source curriculum programme on creating artificial intelligence ethically for children aged 10 to 14. The goal of the programme is to nurture a future of conscientious consumers and developers [Pelzman-Kern, n.d.].

## 5.7 Conclusion

The advent of technology has enabled people to connect and communicate in an unprecedented way. People have taken advantage of certain aspects of technology, such as algorithms, anonymity, and secure platforms, to perpetuate prejudice and emphasise divisions in society. Human tendencies, such as the need to connect with like-minded others, availability heuristics, and confirmation bias, make people inherently vulnerable to prejudiced messages and like-minded people. Some individuals take it further and utilise secure cyber platforms to express their hate in real life. There have been right-wing extremists who started off with cyberhate speech, such as the Westminster attack meme, that eventually escalated to hate crimes offline and even attacks. The interactive nature of the cyber world thus requires online and offline interventions that tackle both the technology and the people who build and use them.

## 5.8 Acknowledgement

The views expressed in this chapter are the author's only and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 5.9 References

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), pp. 1–30.

Allen, E., & Henderson, B. (2017, March 26). Westminster attack: Everything we know so far about the events in London. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/2017/03/22/westminster-terror-attack-everything-know-far/

American Psychological Association. (n.d.). *Availability heuristic.* American Psychological Association. https://dictionary.apa.org/availability-heuristic

Angwin, J., Tobin, A., & Varner, M. (2017, November 21). Facebook (still) letting housing advertisers exclude users by race. *ProPublica.* https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

Anti-Defamation League. (2016). *Responding to cyberhate: Progress and trends.* ADL. https://www.adl.org/sites/default/files/documents/assets/pdf/combating-hate/2016-ADL-Responding-to-Cyberhate-Progress-and-Trends-Report.pdf

Anti-Defamation League. (2018). *Pyramid of Hate*. ADL. https://www.adl.org/sites/default/files/documents/pyramid-of-hate.pdf

Anti-Defamation League. (2010). *Responding to cyberhate: Toolkit for action.* ADL. https://www.adl.org/sites/default/files/documents/assets/pdf/combating-hate/ADL-Responding-to-Cyberhate-Toolkit.pdf

Anti-Defamation League. (n.d.). *Who we are*. ADL. https://www.adl.org/who-we-are

Antinori, A. (2017). From the Islamic State to the 'Islamic State Of Mind': the evolution of the 'Jihadisphere' and the rise of the lone Jihad. *European Law Enforcement Research Bulletin*, (16), pp. 47–55.

Awan, I. (2017). Cyber-extremism: Isis and the power of social media. *Society*, *54*(2), pp. 138–149.

Awan, I., & Zempi, I. (2015). *We fear for our lives: Offline and online experiences of anti-Muslim hostility.* TellMAMA. https://www.tandis.odihr.pl/bitstream/20.500.12389/22288/1/08624.pdf

Bartlett, J., & Birdwell, J. (2013). Cumulative radicalisation between the far-right and Islamist groups in the UK: A review of evidence. *Demos*, *5*. https://pdfs.semanticscholar.org/820f/d4947fecf860a345623d1a7e0802f3858fa6.pdf

Baysinger, T. G. (2006). Right-wing group characteristics and ideology. *Homeland Security Affairs, 2*(2). https://www.hsaj.org/articles/166

Bright, J. (2017). *Explaining the emergence of echo chambers on social media: the role of ideology and extremism.* https://arxiv.org/pdf/1609.05003.pdf

Brown, A. (2018). What is so special about online (as compared to offline) hate speech?. *Ethnicities*, *18*(3), pp. 297–326.

Chang, A. (2018, 2 May). The Facebook and Cambridge Analytica scandal, explained with a simple diagram. *Vox.* https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram

Daud, S. (2019, January 31). S'pore student gets online hate for photo of kiss with male partner, JC 'advised' him to take it down. *Mothership.* https://mothership.sg/2019/01/tmjc-gay-kiss-instagram-hwz/

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific reports*, *6*, 37825.

Dixon, H. (2017, November 13). Russian bot behind false claim Muslim woman ignored victims of Westminster terror attack. *The Telegraph.* https://www.telegraph.co.uk/news/2017/11/13/russian-bot-behind-false-claim-muslim-woman-ignored-victims/

Evolvi, G. (2018). Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses. *Religions*, *9*(10), 307.

Fakhfakh, R., Ammar, A. B., & Amar, C. B. (2017). Deep learning-based recommendation: Current issues and challenges. *International Journal of Advanced Computer Science and Applications, 8*(12), pp. 59–68.

Griffiths, S. (2019, February 11). Can this technology put an end to bullying? *BBC.* https://www.bbc.com/future/article/20190207-how-artificial-intelligence-can-help-stop-bullying

Hafez, F. (2014). Shifting borders: Islamophobia as common ground for building pan-European right- wing unity. *Patterns of Prejudice, 48*(5), pp. 479–499.

Hanzelka, J., & Schmidt, I. (2017). Dynamics of cyber hate in social media: A comparative analysis of anti-Muslim movements in the Czech Republic and Germany. *International Journal of Cyber Criminology*, *11*(1), pp. 143–160.

Hayden, M. E. (2019, June 27). *Far-Right extremists are calling for terrorism on the messaging app Telegram.* Southern Poverty Law Centre. https://www.splcenter.org/hatewatch/2019/06/27/far-right-extremists-are-calling-terrorism-messaging-app-telegram

Hall, S. (2019, May 11). Ukip candidates urge followers to switch to far-right social network Gab. *The Guardian.* https://www.theguardian.com/politics/2019/may/11/ukip-european-election-candidates-join-gab-social-media-far-right

Hern, A. (2017, November 14). How a Russian 'troll soldier' stirred anger after the Westminster attack. *The Guardian.* https://www.theguardian.com/uk-news/2017/nov/14/how-a-russian-troll-soldier-stirred-anger-after-the-west-minster-attack

Hern, A. (2018, January 12). Google's solution to accidental algorithmic racism: ban gorillas. *The Guardian.* https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people

Ho, O., & Aw, C. W. (2016, April 30). PrimDeli sacks staff member over 'racist' remarks. *The Straits Times.* https://www.straitstimes.com/singapore/primadeli-sacks-staff-member-over-racist-remarks

Kurohi, R. (2020, February 8). MHA, Muis investigating religious teacher's posts. *The Straits Times.* https://www.straitstimes.com/singapore/mha-muis-investigating-religious-teachers-posts

L1ght. (2020). *Rising levels of hate speech & online toxicity during this time of crisis.* https://l1ght.com/Toxicity_during_coronavirus_Report-L1ght.pdf

Lim, B. (2019, February 1). 'Is it because I'm Chinese': Go-Jek passenger gets doxxed in wake of viral video. *Asiaone.* https://www.asiaone.com/singapore/it-because-im-chinese-go-jek-passenger-gets-doxxed-wake-viral-video

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, *27*(4), pp. 962–986.

Meredith, S. (2018, March 21). Here's everything you need to know about the Cambridge Analytica scandal. *CNBC.* https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html

Mortimer, C. (2017, November 14). Man who posted image of Muslim woman 'ignoring Westminster terror victims' was a Russian troll. *The Independent.* https://www.independent.co.uk/news/uk/politics/man-muslim-woman-london-terror-attack-phone-russian-troll-identity-a8052961.html

Mudde, C. (2000). The ideology of the extreme right. Oxford: Manchester University Press.

Owen, T. (2019, October 7). How Telegram became White Nationalists' go-to messaging platform. *Vice.* https://www.vice.com/en_us/article/59nk3a/how-telegram-became-white-nationalists-go-to-messaging-platform

Pelzman-Kern, I. (n.d.). *AI + Ethics curriculum for middle school.* MIT Media Lab. https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/

Pulliam-Moore, C. (2015, January 7). Google Photos identified black people as 'gorillas,' but racist software isn't new. *Splinter News.* https://splinternews.com/google-photos-identified-black-people-as-gorillas-but-1793848829

Ravndal, J. A. (2013). Anders Behring Breivik's use of the Internet and social media. *Journal Exit-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur*, *2*, pp. 172–185.

Reid, A. (2016, December 21). Here's how we can protect ourselves from the hidden algorithms that influence our lives. *The Conversation.* https://theconversation.com/heres-how-we-can-protect-ourselves-from-the-hidden-algorithms-that-influence-our-lives-70674

Richterich, A. (2018). How data-driven research fuelled the Cambridge Analytica controversy. *Partecipazione E Conflitto*, *11*(2), pp. 528–543.

Robertson, A. (2019, 10 April). A new bill would force companies to check their algorithms for bias. *The Verge*. https://www.theverge.com/2019/4/10/18304960/congress-algorithmic-accountability-act-wyden-clarke-booker-bill-introduced-house-senate

Roose, K. (2019, August 4). 'Shut the site down,' says the creator of 8chan, a megaphone for gunmen. *The New York Times.* https://www.nytimes.com/2019/08/04/technology/8chan-shooting-manifesto.html

Sarah Camariah. (2016, April 28). So this past Monday … [Facebook status update]. https://www.facebook.com/stillsarahc/posts/10153647453768505?pnref=story

Seaman, J. (2008). Hate Speech and Identity Politics: A Situationalist Proposal. *Fla. St. UL Rev.*, *36*, pp. 99–123.

Shanmuganathan, R. (2019, February 1). Netizens mock woman who accuses Singapore Go-Jek driver tries to kidnap her when he obviously didn't. *The Online Citizen.* https://www.theonlinecitizen.com/2019/02/01/

netizens-mock-woman-who-accuses-singapore-go-jek-driver-tries-to-kidnap-her-when-he-obviously-didnt/

Simonite, T. (2018, November 1). When it comes to Gorillas, Google Photos remains blind. *Wired.* https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

Stalinsky, S., Sosnow, R., Khayat, M., Al-Hadj, M., Green, R., Agron, A., Benjamin, S., & Shemesh, M. (2016, December 23). *Germany-Based Encrypted Messaging App Telegram Emerges as Jihadis' Preferred Communications Platform—Part V of MEMRI Series: Encryption Technology Embraced by ISIS, Al-Qaeda, Other Jihadis—September 2015-September 2016: Section 2—MEMRI Research Documents Jihadi Use Of Telegram.* MEMRI. https://www.memri.org/reports/encryption-technology-embraced-isis-al-qaeda-other-jihadis-part-v-—-september-2015-0

Staples, B. (2018, June 17). The racist trope that won't die. *The New York Times.* https://www.nytimes.com/2018/06/17/opinion/roseanne-racism-blacks-apes.html

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior*, *7*(3), pp. 321–326.

Sunstein, C. R. (2007). Group Polarization and 12 Angry Men. *Negotiation Journal, 23*(4), pp. 443–47.

Tam, L. (2018, September 25). Why stalking, cyberbullying and doxxing are so harmful and what makes people do it. *South China Morning Post.* https://www.scmp.com/lifestyle/family-relationships/article/2165543/why-stalking-cyberbullying-and-doxing-are-so-harmful

Tan, R. (2017, June 30). Terrorists' love for Telegram, explained. *Vox.* https://www.vox.com/world/2017/6/30/15886506/terrorism-isis-telegram-social-media-russia-pavel-durov-twitter

TellMAMA. (2017, March 24). *The truth behind the photo of the Muslim woman on Westminster Bridge.* TellMAMA *UK.* https://tellmamauk.org/the-truth-behind-the-photo-of-the-muslim-woman-on-westminster-bridge/

The Temasek Review (2019, February 2). Jovina Choi … [Status update]. Facebook. https://www.facebook.com/190806675782/posts/jovina-choi-the-passenger-who-falsely-accused-a-go-jek-driver-of-kidnapping-her-/10156597246295783/

Wachs, S., & Wright, M. F. (2018). Associations between bystanders and perpetrators of online hate: The moderating role of toxic online disinhibition. *International Journal of Environmental Research and Public Health*, *15*(9), 2030, https://www.mdpi.com/1660-4601/15/9/2030/pdf

Woodward, C., & Bray, H. (2017, April 4). A company sent anti-abortion ads by phone. Massachusetts wasn't having it. *The Boston Globe.* https://www.bostonglobe.com/business/2017/04/04/healey-halts-digital-ads-targeted-women-reproductive-clinics/AoyPUG8u9hq9bJUAKC5gZN/story.html

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of science, technology & society*, *30*(5), pp. 316–327.

Yong, C. (2020, April 27). Growing wave of racism and hate crimes as coronavirus spreads. *The Straits Times.* https://www.straitstimes.com/world/united-states/growing-wave-of-racism-and-hate-crimes-as-virus-spreads

Zang, J., Dummit, K., Graves, J., Lisker, P., & Sweeney, L. (2015). Who knows what about me? A survey of behind the scenes personal data sharing to third parties by mobile apps. *Technology Science*, *30*. https://techscience.org/a/2015103001/download.pdf

https://doi.org/10.1142/9789811232411_0006

# Chapter 6

# Rebellion Against the State: A Social Perspective on How the Online Space Fuels Collective Action

Hou Minzheng

*National University of Singapore*

*hou.minzheng@u.nus.edu*

## 6.1 Introduction

Since the turn of the century, the world has observed a marked increase in the number of prominent incidents of protests [e.g., Margetts *et al*., 2013; Postmes & Brunsting, 2002]. Many of these incidents involved the galvanisation of, and coordination across hundreds and thousands of people—feats that were hitherto unobserved in history—in pursuit of a collective cause or ideology amidst a political struggle.

Notable examples include the Arab Spring, which started off as a series of anti-government protests against government corruption and autocracy in Tunisia that then led to a wave of uprisings, rebellions, and civil wars across the Arab nations [Blakemore, 2019]; the Hong Kong anti-extradition bill protests, which began in mid-2019 and have since persisted as one of the largest-scale, longest-lasting, and most violent political movements in the history of Hong Kong [Hale & Kollewe, 2019]; and France's yellow vest movement, sparked by rising fuel taxes, which lasted for almost an entire year, and which necessitated the deployment of some 7,500 law enforcement officers for containment efforts [Kaplan & Akhtar, 2019].

The sheer tenacity with which these acts of rebellion have been sustained have led some experts to describe modern society as being in the "Golden Age for collective action" [Margetts *et al*., 2013, p. 279]. Importantly, as evidenced by the scale of these events since the advent of the digital era, researchers have converged at a general consensus on the pivotal role that the online space plays in both the *ignition* and *proliferation* of collective action within and even across geographical borders [e.g., Agarwal, Lim, & Wigand, 2014].

Indeed, a fundamental affordance of the internet lies in its power as a relatively open platform—a medium of mass communication characterised by high accessibility and low barriers to entry [e.g., Postmes & Brunsting, 2002]. As we shall observe in the chapter, this and other defining characteristics of the online sphere shall exert a profound impact on driving collective action against governments.

Against such a backdrop, the current chapter examines how, from a social psychological perspective, the online space can and has served as an enabler for these acts of rebellion. Specifically, we leverage on a dual-pathway perspective of collective action [Sturmer & Simon, 2004] to uncover the role of the Internet in influencing individuals' *identity* and *motivation* in collective action participation.

In line with a social identity perspective [e.g., Tajfel & Turner, 1986; Turner, 1987], I first explain that the online space can ironically present ideal circumstances for the expression of individuals' social, and hence politicised collective identity, thus engendering greater identification with a particular collective cause. *That is, the internet plays an intimate role in fostering a strong politicised collective identity that ignites collective action*.

In addition, I highlight the role of the internet in enhancing activists' motivational drive in initiating and sustaining a movement. That is, *the increased motivational force in participation as afforded by the Internet serves to proliferate collective action*.

Finally, I outline some measures that may help authorities harness the energy of an engaged citizenry towards positive and peaceful national development.

## 6.2 The Role of Identity and Motivation: A Dual-Pathway Perspective

According to a dual-pathway perspective [Sturmer & Simon, 2004], participation in collective action is based on two core tenets: the first of

which pertains to a more affectively based construct of identity, while the second pertains to more cognitively based calculations about the expectations of collective action.

First, collective actions are often ignited through a consciousness of a politicised collective identity [Simon & Klandermans, 2001]. This involves individuals identifying themselves as a group characterised by an *awareness of shared grievances* (e.g., discrimination, unfair treatment, conflicting ideology), *attributions of an adversarial outgroup* (e.g., the government that is perceived as responsible for the grievance), and where individuals engage with one another as group members in a *power struggle against the outgroup* amid the broader context of society.

Following which, the activation of a politicised collective identity inclines individuals to coordinate and take collective action against the adversarial outgroup, depending on one's expectations about the *group's efficacy in engaging in collective action,* as well as the *expected effectiveness of such actions* in achieving the group's collective goals. These expectations influence the motivational force of a group in proliferating collective action.

### 6.2.1 *How the online space ignites politicised collective identity*

According to the Social Identity of Deindividuation Effects (SIDE) [Lea & Spears, 1991; Postmes, Spears, & Lea, 1998], the online space can present ideal circumstances for politicised collective identity to be expressed. Specifically, the theory suggests that the process of depersonalization plays a central role in enhancing the salience of a group's social identity online.

Contrary to early theorising that the online space represents a more individualistic environment, modern social psychologists argue that the "relative anonymity and isolation that characterise many web activities do not always individualise but can also function to enhance group salience by reducing attention to individual differences within the group" [Postmes & Brunsting, 2002, p. 295]. In other words, through the process of depersonalisation which occurs within the relatively anonymous online space, individual differences are pushed to the background while group identity is brought to the fore [Postmes & Brunsting, 2002]. Especially given that social movements or political activism are typically rooted in pre-existing groups (e.g., minorities, LGBTs) [Sturmer & Simon, 2004;

Klandermans, 1984, 1997]—that is, groups interact online on the basis of their shared identities—online groups may thus be united in powerful, psychological ways based on a salient sense of politicised collective identity coupled with a relatively depersonalised sense of self.

Various psychological processes are involved in enhancing the salience of one's social identity online and thus positioning group members on a politicised footing. First, engaging in online communication among group members enhances perceived homogeneity of perspectives based on a "false consensus effect"—the tendency to overestimate public support for one's own perspective [Douglas, 2007; Wojcieszak, 2008, 2011]. Second, the increased perception that group members are interacting with like-minded others strengthens pressure towards conformity to group norms and standards, and thus contribute to the hardening and behavioural enactment of one's politicised collective identity [Alberici & Milesi, 2015]. Indeed, evidence that online norms are just as strong as that offline is testament to the psychological potency of the online environment [Cheung, Chiu, & Lee, 2011].

Facilitated by the online space, the potential for collective action is thus ignited when groups come together online and interact on the basis of an accentuated sense of shared identity. Such a strong consciousness of shared grievances, goals, and intentions result from an important first step towards collective action [Gamson, 1992; Klandermans, 1997].

### 6.2.2 *How the online space proliferates collective action*

Beyond its effect on group member's collective identity, the internet has also served as a critical platform in energising groups towards proliferating collective action. The motivational affordance of the internet may be understood based on the concepts of *efficacy* (i.e., how able are groups in engaging in collective action) and *expectancy* (i.e., how effective collective action is expected to be in achieving the group's collective goals) [e.g., Wigfield & Eccles, 1992; Wigfield & Cambria, 2010; Vroom; 1964].

Specifically, the characteristics of the online space enhances a group's efficacy in engaging in collective action, by enabling coordination across large masses of people as well as influencing public opinion. In addition, the advent of the internet inevitably provides a larger range of means (e.g., online disruptions and petitions, as opposed to limited offline protests and rallies) through which to engage the adversarial outgroup, thus enhancing

the expectancies that collective action can be effective in achieving a group's goals.

### 6.2.3 *Effects of the internet on group efficacy*

To begin, one defining characteristic about the online space is its power of mass communication afforded by its low barriers to entry [Postmes & Brunsting, 2002]. This implies that any online social movement is likely to be observed by a heterogeneous mix of bystanders, sympathisers, or the general public who would not normally be part of the movement.

With such an enhanced communication reach, the online space offers a more efficacious means through which a critical mass may be attained. This is a crucial ingredient in the sustenance of collective action [e.g., Margetts *et al*., 2015]: A critical mass is necessary as part of a typical progression of a mobilisation effort where the existence of sufficient perceived numbers of supporters of a movement enables individuals with more stringent criteria (e.g., greater reservations) to join, thereby strengthening a movement's energy.

In addition, the online space offers a fertile ground for groups attempting to sway public opinion towards their cause. These battles for public opinion remain critical in social movements given that the political solidarity that can be struck between activists and the public is a necessary precondition for any social change to occur [Subašić, Reynolds, & Turner, 2008]. In the historic 2019 Hong Kong protests for example, protesters were prompt in circulating images of a female protester purportedly injured in the eye by police officers, in an effort to depict law enforcement as engaging in excessive brutality [Shao, 2019].

Moreover, whereas traditional forms of offline actions (e.g., protests and strikes) may require effective leadership for social coordination [Calvert, 1992], researchers have argued that with the advent of social media, "formal organisations with structures and incentives are no longer critical" [Bimber, Flanagan, & Stohl, 2012, pp. 4–5]. In this regard, the traditional roles of leaders in cultivating group identity, collective trust, and cohesion may be replaced by the depersonalization process and shared sense of heightened collective identity as described earlier [Margetts *et al*., 2015]. Indeed, by driving individuals towards a stronger commitment to group norms and goals, these processes may explain the rise of many leader-less movements we observe today [Serhan, 2019].

### 6.2.4 *Effects of the internet on expectancies of collective action*

So far, we have examined how various characteristics—increased reach, shifting public opinion, and social identity processes—can contribute to enhanced efficacies of the Internet in fueling collective action. Here, we turn our attention to how the online space can also improve members' expectancies about collective action success.

Specifically, the online environment in itself serves as an alternative means through which collective action goals may be achieved. As with offline behaviour, the online space provides a reservoir of both *persuasive* and *confrontational* action toolkits for group members [Brunsting & Postmes, 2002].

For example, persuasive actions may include online petitions and lobbying, while confrontational actions may include more hostile disobedience elements such as hacking, sit-ins, or blockades [e.g., Wray, 1999]. Such an expanded range of equifinal means (i.e., various approaches leading to the same result) significantly contributes to a group's psychological commitment to their goals [Kruglanski, Pierro, & Sheveland, 2011]. Activists now have more options than ever in driving their message across. The expectations of successfully achieving one's desired outcomes, and thus motivation to pursue these outcomes, may be greatly enhanced.

Taken together, the online space plays a crucial role in the proliferation of collective action. Consistent with the motivational constructs of efficacy and expectancy, the increased reach of the online space can and has been exploited by groups to advance their agenda. The online space has made it easier for activist groups to attain a critical mass for movement sustenance, as well as to shape third-party opinion in the battle for public support. Importantly, the success of engaging in collective action may be less dependent on the existence of a central leadership than ever before and may be achieved through an increased repertoire of available action toolkits afforded by the online space.

## 6.3 Harnessing the Energy of an Engaged Citizenry

Despite the potential for collective action to achieve (positive) social change, acts of protests and strikes—be they online or offline—inevitably disrupt everyday functioning of society. For instance, the 2019 Hong

Kong protests have contributed to a first economic recession in a decade, as well as significant delays in day-to-day legislation proceedings [Leung, 2019].

While we have so far examined how the Internet may be a catalyst for collective action, there is nonetheless room for it to be leveraged to harness the energy of an engaged citizenry, instead of being a source to be condemned by governments. In particular, early detection and appropriate responses to grievances remain key. The objective of such efforts would be to mitigate the polarisation of the struggle between activist groups and the government, in order to prevent an extreme politicised identity to be wilfully enacted.

### 6.3.1 *Online sentiment analysis as a diagnostic tool*

The online environment is a fertile ground replete with the voices of the people. Accordingly, one way in which the Internet may be used to better understand the emergence of grievances is to engage in *sentiment analysis*—the computational treatment of expressions of opinion, sentiment, emotion, beliefs, and speculation in written text [Paltoglou, 2014; Wiebe, Bruce, & O'Hara, 1999].

For example, O'Conner and colleagues [2010] demonstrated the usefulness of sentiment analysis techniques in predicting political opinion and consumer confidence; Mishne and de Rijke [2006] provided evidence of changes in public emotions displayed through blogposts in response to both recurring and irregular events. These findings attest to the robustness of sentiment analysis techniques in tracking evolving public sentiments and attitudes towards various events and targets. As such, authorities may choose to implement sentiment analysis as one of the mainstay research methodologies to detect sentiments of concern voiced by various social groups.

### 6.3.2 *Surveys and dialogues as diagnostic tools*

Nevertheless, given demographic asymmetries in online participation [Albrecht, 2006], it may be necessary to engage in efforts beyond sentiment analysis in order to obtain a more representative and accurate assessment of actual ground sentiments.

Turning our attention to efforts in Singapore as an example, nation-wide surveys such as those conducted by the Institute of Policy Studies

[IPS, 2015] remain a vital means of engaging the population and identifying emerging trends. For instance, according to IPS' Perception of Policies in Singapore Survey [2015], issues relating to fairness of government policy, need for checks and balances, need for different views in parliament, as well as cost of living are significant areas that voters are concerned with. In addition, the survey revealed that the proportion of 'pluralist' voters (i.e., those whose responses suggest support for greater mix of political voices and changes in the electoral system) increased between 2006 to 2011, but decreased between 2011 to 2015, representing a change in voter sentiments towards political pluralism according to evolving sociopolitical climate. Establishing an understanding of various demographic profiles (e.g., based on education, socioeconomic status etc.) that constitute political 'pluralists' may therefore permit a more targeted engagement approach in addressing concerns about governance.

Other means to understand public sentiments include quick, regular polls on key areas to track fluctuations in sentiments across time. The conduct of monthly polls by Blackbox Research on a suite of areas constituting government satisfaction (e.g., public transport, management of economy, racial relations) is an example ["Government satisfaction index," 2019]. Close monitoring of sentiments of Singaporeans towards significant global events (e.g., regional protests in Hong Kong and Indonesia, global climate change movements, human rights movements such as LGBTQ issues) also provides an important source of information to understand how our people respond to these influential occurrences. Such close monitoring efforts may be done via online sentiment analysis of selected social media groups, straw polls by major newspapers or research agencies, government feedback units such as REACH, and observing the nature of discussions at various professional or academic forums in terms of people's prevalent concerns.

Beyond the above measures that captures the breadth of public sentiment evaluation, it is also important to engage in dialogues with citizens to achieve depth in understanding ground issues. These dialogues can serve both as a diagnostic tool to assess ground sentiments, as well as to demonstrate empathy and genuine concern towards grievances. Such efforts, including the SGfuture dialogues [Heng, 2019], that emphasise building the future society together, are paramount in demonstrating a responsive government and preventing a fractured society.

### 6.3.3 *Clear and transparent communication channels as intervention tools*

Last but not least, it is important for government agencies to ensure that communication channels—online and offline—with the public remain *clear* (i.e., people know where to go to officially register their concerns) and *transparent* (i.e., people trust that their concerns are registered and responded to).

Encouraging citizens to leverage on these channels not only enable agencies to take prompt action to address any emerging issues, but also to prevent discontent from festering and tipping over into politicised grievances. It would be most undesirable if the only way people can have their voices heard is by organising collective action—online or offline—to pressure government agencies into complying.

### 6.3.4 *A final note on interventions*

Any effective intervention must be able to address the associated grievance at its root. Specifically, in the domain of collective action against the State, this inevitably means actual policy changes to address the grievance, and/or communication efforts to help citizens understand the rationale of various policies.

In other words, any effective intervention ultimately boils down to *good governance*—the ability to unite society and avoid the marginalisation of any social group. In this vein, the above recommendations are aimed at assisting government agencies to detect and understand the *what* and *whys* of grievance promptly, which serve as the basis for governments' formulation of appropriate policy changes and/or communication efforts.

## 6.4 Conclusion

In this chapter, I have sought to describe from a social psychological perspective how the Internet has shaped social identity processes and core motivational constructs in the ignition and proliferation of collective action. The online space facilitates the accentuation of one's politicised collective identity and promotes increased behavioural intentions according to shared norms and goals of the collective group. In addition,

the Internet can enhance the efficacies related to achieving a critical mass necessary for a movement's sustenance, shaping critical public opinion, as well as enable leader-less forms of coordination to emerge. The suite of online disruptions available at the hands of would-be activists further lend viability to collective action as a desirable option.

Despite the enabling factors of the online space in driving collective action, there are nevertheless areas in which the Internet may be exploited for the early detection of emerging trends. Through online sentiment analysis, surveys and dialogues, governments may engage with diverse groups in society in a timely manner, and thus mitigate the potentially disruptive nature of collective action. Communication channels must also remain open, and genuine effort must be taken to implement necessary policy changes and/or communicate policy rationales. Ultimately, these measures help harness the energy of an engaged citizenry and prevent the fragmentation of society.

## 6.5 Acknowledgement

The views expressed in this chapter are the author's alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 6.6 References

Agarwal, N., Lim, M., & Wigand, R. T. (2014). *Online collective action: Dynamics of the crowd in social media* (2014th ed.). Wien: Springer. doi:10.1007/978-3-7091-1340-0

Alberici, A. I., & Milesi, P. (2016). Online discussion, politicized identity, and collective action. *Group Processes & Intergroup Relations*, *19*(1), pp. 43–59. doi:10.1177/1368430215581430

Albrecht, S. (2006). Whose voice is heard in online deliberation?: A study of participation and representation in political debates on the internet. *Information, Communication & Society*, *9*(1), pp. 62–82. doi:10.1080/1369-1180500519548

Bimber, B., Flanigan, A., & Stohl, C. (2012). *Collective Action in Organizations*. Cambridge: Cambridge University Press.

Blakemore, E. (2019). What was the Arab Spring and how did it spread?. *National Geographic*. Retrieved from: https://www.nationalgeographic.com/culture/topics/reference/arab-spring-cause/

Brunsting, S., & Postmes, T. (2002). Social movement participation in the digital age: Predicting offline and online collective action. *Small Group Research*, *33*(5), pp. 525–554. doi:10.1177/104649602237169

Calvert, R. (1992). 'Leadership and its Basis in Problems of Social Coordination', *International Political Science Review*, *13*, pp. 7–24.

Cheung, C. M. K., Chiu, P., & Lee, M. K. O. (2011). Online social networks: Why do students use facebook? *Computers in Human Behavior*, *27*(4), pp. 1,337–1,343. doi:10.1016/j.chb.2010.07.028

Douglas, K. M. (2007). Psychology, discrimination and hate groups online. In A. N. Joinson, K. Y. McKenna, T. Postmes, & U. D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 155–163). Oxford, UK: Oxford University Press.

Gamson, W. A. (1992). The social psychology of collective action. In A. D. Morris & C. M. Mueller (Eds.), *Frontiers in social movement theory* (pp. 53–76). New Haven, CT: Yale University Press.

Government satisfaction index. (September 2019). *Blackbox Research*. Retrieved from: https://www.blackbox.com.sg/2019/10/18/government-satisfaction-index-gsi-september-2019/

Hale, E., & Kollewe, J. (2019, August 11). Hong Kong hit by more violence as protests enter 10th week. *The Guardian*. Retrieved from: https://www.theguardian.com/world/2019/aug/11/hong-kong-protesters-play-cat-and-mouse-game-with-police-on-subway

Heng, S. K. (2019). Speech by DPM and Minister for Finance Heng Swee Keat at the "Building Our Future Singapore Together" dialogue on 15 June 2019. Retrieved from: https://www.pmo.gov.sg/Newsroom/DPM-Heng-Swee-Keat-Building-Our-Future-Singapore-Together-Dialogue

Institute of Policy Studies (2015). POPS (8)—IPS Post-election survey 2015. Retrieved from: https://lkyspp.nus.edu.sg/ips/research/surveys/completed-surveys/ips-perception-of-policies-in-singapore-(pops)-survey/pops-(8)-ips-post-election-survey-2015

Kaplan, J., & Akhtar, A. (2019, October, 22). A world on fire: Here are all the major protests happening around the globe right now. *Business Insider Singapore*. Retrieved from: https://www.businessinsider.sg/all-the-protests-around-the-world-right-now/?r=US&IR=T

Klandermans, B. (1984). Mobilization and participation: Social psychological expansions of resource mobilization theory. *American Sociological Review*, *49*, pp. 583–600.

Klandermans, B. (1997). *The social psychology of protest*. Oxford: Blackwell.

Kruglanski, A. W., Pierro, A., & Sheveland, A. (2011). How many roads lead to rome? Equifinality set-size and commitment to goals and means. *European Journal of Social Psychology*, *41*(3), pp. 344–352. doi:10.1002/ejsp.780

Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man Machine Studies*, *34*, pp. 283–301.

Leung, K. (2019, December 7). Hong Kong protests have cost government billions in tax dollars after trashing of Legislative Council building in July. *South China Morning Post*. Retrieved from: https://www.scmp.com/news/hong-kong/hong-kong-economy/article/3041082/hong-kong-protesters-have-cost-government-billions

Louis, W. R. (2009). Collective Action—and then what? *Journal of Social Issues*, *65*(4), pp. 727–748. doi:10.1111/j.1540-4560.2009.01623.x

Margetts, H. Z., John, P., Hale, S. A., & Reissfelder, S. (2015). Leadership without leaders? starters and followers in online collective action. *Political Studies*, 63(2), pp. 278–299. doi:10.1111/1467-9248.12075

Mishne, G., & de Rijke, M. (2006). Capturing global mood levels using blog posts. In: *Proceedings of AAAI-CAAW*, Stanford University, Stanford, CA, pp. 145–152

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In: *Proceedings of ICWSM'10*, Washington, DC.

Postmes, T., & Brunsting, S. (2002). Collective action in the age of the internet: Mass communication and online mobilization. *Social Science Computer Review*, 20(3), pp. 290–301. doi:10.1177/089443930202000306

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer mediated communication. *Communication Research*, 25, pp. 689–715.

Paltoglou, G. (2014). Sentiment analysis in social media. In Agarwal, N., Lim, M., & Wigand, R. T. (Eds.), *Online collective action: Dynamics of the crowd in social media* (2014th ed.). Wien: Springer. doi:10.1007/978-3-7091-1340-0

Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 6, pp. 161–198). Chichester, UK: Wiley.

Serhan, Y. (2019, November 19). The common element uniting worldwide protests. *The Atlantic*. Retrieved from: https://www.theatlantic.com/international/archive/2019/11/leaderless-protests-around-world/602194/

Sim, W. (2016, January 23). Time for citizens to start ground-up national conversations. *The Straits Times*. Retrieved from: https://www.straitstimes.com/politics/time-for-citizens-to-start-ground-up-national-conversations

Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American Psychologist*, *56*(4), pp. 319–331. doi:10.1037/0003-066X.56.4.319

Sturmer, S., & Simon, B. (2004). Collective action: Towards a dual-pathway model. *European Review of Social Psychology*, *15*(1), pp. 59–99. doi:10.1080/10463280340000117

Subašić, E., Reynolds, K. J., & Turner, J. C. (2008). The political solidarity model of social change: Dynamics of self-categorization in intergroup power relations. *Personality and Social Psychology Review*, *12*(4), pp. 330–352. doi:10.1177/1088868308323223

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *The psychology of intergroup relations* (2nd ed., pp. 7–24). Chicago: Nelson-Hall.

Tuomela, R. (1995). *The importance of us: A philosophy study of basic social notions*. Stanford, CA: Stanford University Press.

Turner, J. C. (1987). A self-categorization theory. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the social group*, (pp. 42–67). Oxford, UK: Basil Blackwell.

Vroom, V. H. (1964). *Work and motivation*. Oxford, UK: Wiley.

Wiebe, J., Bruce, R. F., & O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In: *Proceedings of 37th annual meeting of ACL*, College Park, MD, USA, pp. 246–253.

Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. In S. Karabenick & T. Urdan (Eds.), *Advances in Motivation and Achievement*. The Next Decade of Research on Motivation (Vol. 16). New York: Taylor Francis Group

Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*, pp. 265–310.

Wojcieszak, M. E. (2008). False consensus goes online: Impact of ideologically homogeneous groups on false consensus. *Public Opinion Quarterly*, *72*, pp. 781–791. doi:10.1093/poq/nfn056

Wojcieszak, M. E. (2011). Computer-mediated false consensus: Radical online groups, social networks and news media. *Mass Communication & Society*, *14*, pp. 527–546. doi:10.1080/15205436.2010.513795

Wray, S. (1999). On electronic civil disobedience. *Peace Review*, *11*(1), pp. 107–111. doi:10.1080/10402659908426237

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0007

# Chapter 7

# Victim and the Cyber Vigilante: An Additional Perspective on Cyber Vigilantism

Yasmine Wong

*Centre of Excellence of National Security,*
*S. Rajaratnam School of International Studies,*
*Nanyang Technological University*

*isyasminewong@ntu.edu.sg*

## 7.1 Introduction

Vigilantism is an enigmatic concept, it is, at its core, the regulation of criminal deviance through informal and illegitimate means [Johnston, 1996]. That it has resulted in the exaction of social justice in some instances (real or perceived): despite its questionable legality, has contributed to its appeal in many societies.

With the emergence, and increased accessibility to new communication technologies, the individual's capacity to contribute to and participate in online discussions has been augmented [Castells, 2008]. With the advent of Web 2.0, this new freedom to create and share content online has empowered individuals to exact justice online, giving rise to a new kind of vigilantism. Netizens can now unearth, or gain access to, previously obscure, anonymous, or protected personal data of wrongdoers from their social media accounts and/or other online platforms [Cheong & Gong, 2010]. This has given rise to online shaming and 'doxxing', with the latter

defined as the sharing of personal details (e.g., name, home address, phone number, pictures) of the targeted individual with the public, by publishing them on a public website [Douglas, 2016]. This new vigilantism—cyber vigilantism—can be understood as a planned act carried out by private autonomous individuals, or online communities formed in response to a perceived violation of social norms; with the aim of bringing to light perceived transgressions identified by netizens online, and to eventually carry out acts of punishment against the perceived perpetrator with the purported aim of providing assurances of safety for the online community [Wong & Low, 2019].

Expectedly, experts say that cyber vigilantism will only become more commonplace with the proliferation and accessibility of recording devices and social media [Lin, 2017]. The intended and unintended consequences of this trend have been damaging, and as such, some countries have adopted measures to deal with the phenomenon. In Singapore, for example, amendments were made to the Protection from Harassment Act (POHA) in 2019, which now criminalises doxxing [Ministry of Law, 2019]. Despite the changes in the law, netizens continue to engage in online shaming and doxxing. For example, during the COVID-19 pandemic, instances of cyber vigilantism were seemingly on the rise as netizens posted videos and some cases, doxxed individuals who flouted safe distancing rules on social media pages like 'SG Covidiot' [Pan, 2020].

Cyber vigilantism expectedly poses various challenges in the local context. This chapter will focus on two of them. One, the use of doxxing and virtue online shaming by victims of sexual harassment and assault against their perpetrators. Two, in this context, why individuals unrelated to the victim would resort to doxxing even though they could be breaking the law.

This chapter will first examine why victims of sexual harassment and assault might use doxxing against their perpetrators. Secondly, it explores the cascading mobilisation of the public in this context, which could lead to collective participation in cyber vigilante behaviours. Finally, the chapter concludes with some points for further consideration.

## 7.2 Use of Cyber Vigilantism by Victims of Sexual Harassment and Assault

On the 18 April 2019, netizens in Singapore were made aware of the alleged mishandling of a sexual harassment case at the National University

of Singapore (NUS), where a student was caught filming another student in the shower at a hostel on campus [Ng & Choo, 2019]. The novelty of this case was how it entered the sphere of public consciousness—the victim, Ms. Monica Baey, released a series of Instagram stories detailing the incident, what she considered to be inadequate punishment meted out to the perpetrator, as well as revealing the personal information of the perpetrator [Ang, 2019]. In what some referred to as a "trial by social media," the perpetrator was harassed and shamed online, while the Singapore Police Force (SPF) and legal system were severely criticised by netizens.

Baey's decision to speak of her experience online is not an anomaly in Singapore and stands amid others who have done the same. Ms. Karmen Siew went on Facebook to express her disappointment in the way her case was handled by the criminal justice system, perceiving her perpetrator to have received an insufficient sentence [Cheow, 2019]. A Temasek Polytechnic student posted a picture of a man (withholding his face and identity) being questioned by two police officers inside a toilet [Lim, 2019]. She described in the caption that she and a friend were in the toilet when the man tried to force the door open [Lim, 2019]. She decided to share her story online to ensure coverage of her experience to serve as a warning for others that such incidents do happen [Lim, 2019]. In the same year, police investigated reports against the SG Nasi Lemak chat group, in which users shared pictures of young women and girls, to which other members would often respond with lewd comments [Lim, 2019]. When the pictures of one woman was circulated in the chat by a man whom she had known as a friend, the woman took to Instagram and posted the man's pictures, name, and other identifying details to recount what he had done [Lim, 2019].

### 7.2.1 *International influences*

Their decision to share their stories online has made salient the issue of sexual harassment and assault in Singapore, sparking discussions on what can be done to tackle an issue that is often sensitive [Dutta, 2019]. This mirrors the individuals who spoke out against sexual abuse and harassment and rallied under the #MeToo banner. Named Person of the Year 2017 by TIME magazine, the 'silence breakers'—women and men who spoke out against sexual abuse and harassment—are representative of the #MeToo

movement which sprung up as allegations emerged against Hollywood producer Harvey Weinstein ["Person of the Year: Time," 2017]. The movement and its hashtag were founded in 2006 to encourage survivors of sexual violence to speak up about their experiences and find pathways to healing [metoomvmnt.org, n.d.]. The movement took off on social media and became a global sensation, with women, as well as men, flooding social media with their stories of being harassed using the #MeToo hashtag [Langone, 2018]. These narratives often highlighted a common theme—speaking out against people in power, or against a system that traps victims in unfavourable positions of power, with the intention to hold perpetrators accountable and to enact systemic changes. Since the start of the #MeToo movement, the Association of Women for Action and Research's (AWARE) Sexual Assault Care Centre (SACC) has reported the doubling of calls from 35 to 70 cases per month [Amour-Levar, 2017]. This suggests the #MeToo movement has gained traction locally, emboldening victims to seek help [Amour-Levar, 2017].

In addition, the use of cyber vigilantism by victims of sexual harassment and assault arguably coincides with the rise of a cultural phenomenon made popular by social media—cancel culture. Cancel culture is a term that typically describes the idea of the public 'blocking' someone who has done something offensive, either through what is essentially a boycott, or attempting to get them disciplined for their actions [Romano, 2019]. Cancel culture features strongly in pop culture, with a long list of celebrities from R. Kelly being cancelled for sexual predation to Kevin Hart being cancelled after netizens unearthed homophobic and racist jokes he has made in the past [Romano, 2019]. Although typically used in the context of the public preventing a public figure from having further prominence or career in the public sphere [Romano, 2019], the online debate and public backlash that follows is a phenomenon that can be widely observed on social media beyond the sphere of celebrity boycott. The characteristic naming and denouncing follow the common cyber vigilante practice of doxxing and shaming; both emphasising the role of people power in meting out justice. The influence of cancel culture can be seen most prominently on social media, with the adoption of the term "cancel" by netizens when discussing what they consider to be problematic. Cancel culture has been regarded as an extension of social justice and the push for equality—a long overdue means of holding power accountable [Romano, 2019]. It is also perceived to be an effective tool to achieve justice online, as the social rewards

cancel culture produces are immediate and gratifying, whilst at the same time distancing the participant from danger [Henderson, 2019]. Furthermore, when people participate in cancel culture, they are showing solidarity with others who share the same viewpoint, likewise, they enjoy the solidarity shared by others, thus strengthening the social bonds of the collective that comes together against a perpetrator [Henderson, 2019].

However, despite being driven by a desire to hold problematic individuals accountable for their actions, cancel culture risks encouraging the adoption of a binary tribal mentality, as in cases of cyber vigilantism. It encourages the public to reduce the issue at hand to simple good versus evil, to view the subjects purely as victim versus abuser [Brooks, 2019]. Binary thinking may create boundaries between groups of people [Robbins, 2015], in this case, between the victim and perpetrator. This narrows the avenues for apology, forgiveness and mediation [Brooks, 2019], which are channels that traditional legal processes aim to provide.

### 7.2.2 *Cyber vigilantism and empowerment of victims of sexual harassment and assault*

The use of cyber vigilantism by victims can also be understood as a form of empowerment. Empowerment focuses on the idea of genuine opportunity for victims to participate in and influence the criminal justice response [Van Ness & Strong, 2006]. From a psychological perspective, empowerment involves personal perception of control (self-efficacy) and involvement in community activities [Chang & Poon, 2017]. Zimmerman [2000] also argues that citizens who have engaged in community activities perceive their levels of competence and control to be higher, reinforcing the idea of participation as fundamental to empowerment. Social media as an empowering medium allows victims to choose between the available alternatives to resolve one's own matter [Barton, 2000]. It is often noted that such technologies have been adopted by marginalised individuals and communities to aid in the advancement of various political and civic agendas [Soon & Cho, 2013]. In this case, victims often speak up on social media, sometimes even doxxing their perpetrators. Often, they are motivated by the belief that their cases were mishandled and to shed light on the perceived structural flaws in the criminal justice system and society at large [Brown, 1975]. This is because the participatory nature of social

media allows users to gain exposure to a greater audience as well as the platform to produce their own narratives.

The access to an audience facilitated by social media is, in itself, empowering. Power, which can be hard to achieve in the real world, is often the motivator for participation in cyber vigilantism [Chang & Poon, 2017], as social media provides an avenue through which victims are given the power to produce and control the narratives surrounding their circumstance. Borrowing from a Foucauldian understanding of power, Castells [2011] suggests that to challenge the existing power relations in society, it is necessary to produce alternative discourses that have the potential to overwhelm dominant narratives. As such, social media enables victims to usurp traditional power structures through the telling of a personal narrative to members of a support network. Thus, allowing them to break the silence that victims are commonly trapped in [Montalbano, 1995].

## 7.3 Use of Cyber Vigilantism by Other Individuals in the Context of Sexual Harassment and Assault

The collision of individual empowerment with international movements that reinforce the significance of speaking out against injustices creates a powerful impetus for action. These factors suggest that sharing experiences on social media not only empowers victims but can also serve to rally collective action. The media has always played a pivotal role in the organisation of social movements, and the rise of social media has further revolutionised the way these movements function [Soon & Cho, 2014]. Social media allows for individuals to be producers of information, unlike traditional media which broadcast dominant narratives of those in power. This allows for disparate peoples to organise themselves, leading to social empowerment through participation [Fenton, 2016]. Zimmerman *et al*.'s [1992] conception of empowerment includes a behavioural element, in which individuals engage in behaviours to help others to cope, often organising others who share similar concerns. Social media therefore makes it possible for people who share common grievances to converge with ease and speed [Chang & Poon, 2017]. Participation is also enhanced on social media, in which messages are channelled through

dense social networks, allowing people to respond and share their own stories and concerns [Bennett, 2012]. This prevalent use of social media allows individuals to tap onto their own social networks, thus becoming important catalysts for collective action [Bennett, 2012].

Social media not only allows individuals with common grievances to form collectives, it also allows for the involvement of the public—a public which may not share the same affliction, but one that sympathises with the cause of the collective. This is sometimes described as "mob justice" [Jagdish, 2018], where the court of public opinion serves as judge and executioner. When netizens learn of an injustice from social media, they have the option to engage in discussions surrounding that issue, 'experiencing' and validating one another's emotional responses. This creates an environment where the emotions of one individual can easily 'trigger' similar emotions in another [i.e., emotional contagion; Wong & Loh, 2019]. As such, the sharing of a personal grievance may trigger netizens to empathise with the emotional trauma of the victim, especially so when the global reach of the #MeToo movement suggests a systemic problem. This leads to the cascading mobilisation of support against the perpetrator, as netizens participate in doxxing and shaming behaviours in an attempt to right the injustice.

Indeed, a survey conduct with 1,000 Singaporeans shows that although 95% of respondents are in favour of the criminalisation of doxxing, 65% of this number reflect that some exceptions should be made [Most Singaporeans Support the Criminalisation, 2019]. In the same survey, 31% of respondents state that they would doxx someone if that person has carried out an illegal/undesirable act [Most Singaporeans Support the Criminalisation, 2019]. When different case studies were tested, the difference in public opinion point to the idea that public dissatisfaction at the perceived lack of severity of a punishment can contribute to the justification of doxxing [Most Singaporeans Support the Criminalisation, 2019]. The outpouring of public dissatisfaction with regard to the sentencing of sexual offenders in Singapore which includes petitions against the sentences [Sim, 2019; Teng, 2019a] are indicative of public perception of insufficient punishment. As such, although legislation criminalising doxxing is well-received in Singapore, support for the act itself still remains in situations where doxxing is deemed necessary to promote more just results.

## 7.4 Potential Concerns Arising from the Adoption of Cyber Vigilante Behaviours

Victims who resort to and/or utilise cyber vigilantism against their perpetrators expectedly create challenges as well. First, their usage of cyber vigilantism may be illegal in their jurisdictions, even though they may perceive the usage of such an approach as justifiable. In countries like Singapore and the UK, doxxing regardless of the motivation is an offence [Neo, 2019; "Internet trolls targeted with new," 2016]. The complexity of the situation is compounded when individuals believe the offence (cyber vigilantism) is an appropriate response and aid the victim in perpetuating the offence and/or encourage the victim to perpetrate the offence. The Temasek Polytechnic student who posted a picture of the man who tried to barge into her toilet cubicle, for example, garnered over 21,000 likes on the image she posted on her Instagram, with many netizens leaving words of comfort for her and angry comments for the perpetrator [Lim, 2019]. Baey similarly received support from netizens for putting her attacker's personal information online, and the subsequent acts of doxxing and shaming carried out by netizens against the perpetrator were often goaded by other netizens [Wong & Loh, 2019].

If individuals are allowed to exact 'justice' through non-legal means, even if it occurs in cyberspace, it sends a signal that the rule of law, due legal processes, and sentencing (i.e., ensuring sentences commensurate with the culpability of the offender and the severity of the crime) may be ineffective [Thian, 2014]. This may encourage more individuals to turn to cyber vigilantism rather than legal methods as a recourse.

Third, notwithstanding the motivations for resorting to cyber vigilantism, using such methods can encourage more cyber vigilante acts, and in some cases, more extreme forms of cyber vigilante acts including violence [Online Hate Prevention Institute, 2015]. Zetter [2007] argues that acts of cyber vigilantism exacerbate anarchy online because activity in cyberspace is difficult to regulate. Furthermore, cyberspace is an environment where a lot of information is available, and in such an environment, individuals tend to look for "popularity cues" in order to assess and decipher their experiences online [Fu & Sim, 2011]. As such, individuals may adopt the view of the majority, under the assumption that the majority is usually right [Nadeau, Cloutier, & Guay, 1993]. Thus, public support for cyber vigilantism, influenced by the increasing cultural

prominence of cancel culture exacerbates the situation, contributing to the normalisation of such behaviours.

Fourth, increases in cyber vigilante witch-hunts often lead to psychological damage to all parties targeted [Tam, 2018]. This is especially so for individuals who are wrongly targeted as a result of mistaken identity ["As the 'human flesh search engine'," 2019]. For example, a Canadian Sikh was mis-identified as the individual responsible for the 2015 terror attack in France. He was harassed and received death threats, and suffered immense emotional trauma as a result [Jubbal, 2016].

## 7.5 Points for Further Thought and Deliberation

The desire to participate in the promotion of social justice often results in cascading participation in cyber vigilantism. However, this has led to the blurring of lines between raising social awareness of injustices and online shaming [Tang, 2020]. The use of cyber vigilante behaviours by victims and the collective action that follows may throw into question what is considered deviant behaviour. The following issues require further thought and deliberation.

### 7.5.1 *Accounting for victims and victim empowerment*

It would be overly simplistic to label victim participation online as deviant. Equally, there has been debate on whether cancel culture can be attributed to mob mentality, or if it is a meaningful way of speaking up against problematic power structures [Romano, 2019]. As such, to equate participation through social media as mere mob mentality, would disregard the genuine desire for participation and change' as well as the larger public's participation through social media as mere mob mentality, would disregard genuine desire for participation and change. Following the concept of empowerment, empowered individuals should be understood as citizens with rights and means, and not as individuals dependent on traditional services provided by the state and relevant authorities [Rappaport, 1981].

Besides regulating activities on the internet and curbing deviant activities, there is a need for law enforcement and public policy to recognise the reality that the internet has made it conducive for the public to act on their desires to enact justice [Nhan, Huey, & Broll, 2017],

at times through means typically considered deviant. There is a need to situate the reality of victim empowerment in the current legal context as the prevalence of social media has resulted in the pervasive transformation of how individuals behave and the means through which they are able to act as a collective. Therefore, victims should be actively managed, by providing them with timely resources (e.g., helplines, support groups). An example of this resource in Singapore is the Victim Care Cadre[a] which offers support to trauma victims, mostly of sexual offences, as part of police investigations [Aw, 2018]. These volunteers provide information on processes and how to cope with trauma and can also be employed to inform victims of the doxxing laws.

In the same vein, there were also conversations surrounding victim compensation during the tabling of the revision to the Singapore's Penal Code [Vijayan, 2018]. The proposed changes included measures that would allow victims to participate directly in the compensation process in court [Vijayan, 2018]. Perhaps this can be explored to enhance avenues through which victim narratives and experiences can be incorporated into the criminal justice processes. Such initiatives may restore victims' trust in law enforcement processes as victims often engage in vigilante behaviours as an informal way to offset the inadequacies of the formal justice system [Burrows, 1976].

Besides updating the laws to account for new technologies [Wong, 2019], it is similarly pertinent for Singapore to review old laws to ensure the safety and protection of its people. A good case in point is the Criminal Law Reform Bill, where marital immunity for rape has been repealed and the definition of rape has been expanded beyond vaginal penetration [Kwang, 2019]. This redefinition of rape reflects the desire to expand and strengthen protection against sexual abuse, accounting for a more nuanced understanding of vulnerability. Criminal law embodies the norms and values of society [Tomer-Fishman, 2010], and tightening laws against certain crimes can serve to send a message of societal intolerance for such behaviours. It is also paramount that institutions review internal processes for handling sexual offences, as local universities have done in the wake of fierce debate [Teng, 2019b].

---

[a]The Victim Care Cadre (VCC) Programme provides training for volunteers to provide emotional support to victims, as well as specialist courses for investigation officers handling sexual crime cases (Mokhtar, 2017).

### 7.5.2 *Public perception of rehabilitative practices*

As Singapore is moving to allow greater flexibility in sentencing with more graduated sentencing options for some offences [Koh, 2016], it would be helpful for us to better understand public perceptions of justice. This is because the public can be motivated to engage in cyber vigilante behaviours by the desire to preserve their perception of whether justice has been served. When presented hypothetical situations that describe crimes of differing severity, individuals would often mete out punishments that are perceived to be proportional to the severity of the crime [Darley, Carlsmith, & Robinson, 2000], thus reflecting a 'just deserts' perspective to crime and punishment. However, the criminal justice system exists to fulfil functions beyond punishment, placing emphasis on rehabilitation as well [Ganapathy, 2018]. The expansion of community sentencing, which is the combination of punishment with rehabilitation aligns with this.

Statistically, Singapore has seen success in reducing its criminal reoffence rate [Khair, 2018], and the importance of this success should be communicated to the public as a reflection of its success in rehabilitating offenders. Thus, this also highlights the role of initiatives like The Yellow Ribbon Project, and the pivotal role of rehabilitation in keeping recidivism rates low. Such narratives should be enhanced and continued to reduce the 'just deserts' perspective on law enforcement and recalibrate the public's understanding of justice beyond punishment.

### 7.5.3 *Channelling public participation*

Furthermore, despite the acknowledgement of some of its unsavoury consequences, crowdsourced participation can be harnessed for good [Wong & Loh, 2019]. The sustained public support of the act despite its criminalisation points to the need to consider the public's desire to promote just outcomes as a strong motivating factor. As such, the provision of avenues to direct this behaviour in a constructive manner may be a constructive way to curb the expression of social justice through the means of doxxing and shaming.

Additionally, in cybersecurity literature, multistakeholderism is essential for the holistic management of cyberspace, grounded on the fact that the internet permeates public and private infrastructures managed by diverse actors [Silva, 2017]. Citizen participation can serve as extra sets

of 'eyes and ears' for law enforcement [Nhan, Huey, & Broll, 2017], especially on the internet, providing valuable information for relevant agencies to act on. To diminish shaming and punishing behaviours, the authorities could moderate the process of crowdsourced investigation by sieving out shaming comments, promoting healthier discourse [Wong & Loh, 2019], and collaborating with netizens. A positive example would be the way in which the Canadian authorities handled the Vancouver Stanley Cup Riot in 2011, where the police created a secure webpage for people to submit pictures of the riot and actively sought the help of the public to identify perpetrators in the photos [McCann, 2011]. This exemplifies the potential security capital of the public when mobilised appropriately [Nhan, Huey, & Broll, 2017], reflecting the potential for civic participation in law enforcement. In Singapore, this can be done through the promotion of official platforms such as the i-Witness platform created by the Singapore Police Force.

## 7.6 Conclusion

The debate surrounding victims and their decision to reveal the identity of their perpetrators can be polarising. On one hand, lies repercussions in the field of data protection [Silva, 2017] and the infringement of the rule of law. On the other, lies the desire to uphold a certain conception of justice. As society aims to better regulate the internet and social media, laws against deviant behaviours online must consider a more nuanced understanding of cyber vigilantism and the reasons behind its strong public support. The employment of social media by victims to share their stories has reproduced the symbolism of social media as spaces in which marginalised voices can be heard. The public support that follow may pose challenges to traditional processes and structures when the group turns to cyber vigilantism to mete our justice. As such, laws criminalising a limited definition of cyber vigilante behaviours can be complemented with other measures to regulate such behaviours online—measures that capitalise on the empowerment of individuals and collectives towards the betterment of society.

## 7.7 Acknowledgement

The views expressed in this chapter are the author's own and do not represent the official position or view of their organisational affiliations.

## 7.8 References

Aertsen, I., Bolivar, D., De Mesmaecker, V., & Lauwers, N. (2011). Restorative justice and the active victim: Exploring the concept of empowerment. Temida, pp. 5–19.

Ang, M. (2019). NUS student who revealed male perpetrator's details could be breaking proposed doxxing laws. Mothership. Retrieved from https://mothership.sg/2019/04/nus-student-doxxing-peeping-tom-harassment/

Amour-Levar, C. (2017). The #MeToo movement in Asia: Is Singapore feeling the Weinstein effect?. Forbes. Retrieved from https://www.forbes.com/sites/christineamourlevar/2017/12/17/sexual-harassment-in-the-workplace-is-singapore-feeling-the-weinstein-effect/#118ca245160d

Aw, C. W. (2018). Being there for the victims of trauma. Straits Times. Retrieved from https://www.straitstimes.com/singapore/being-there-for-the-victims-of-trauma

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), pp. 644–675.

Barton, C. (2000). Empowerment and retribution in criminal justice. In H. Strang & J. Braithwaite (Eds.), Restorative justice: Philosophy to practice. Aldershot: Ashgate/Dartmouth, pp. 55–76.

BBC. (2016). Internet trolls targeted with new legal guidelines. BBC. Retrieved from https://www.bbc.com/news/uk-37601431

BBC (2017). Person of the Year: Time honours abuse 'silence breakers', BBC, December 6th [online], accessed on 9 April 2018. Retrieved from http://www.bbc.co.uk/news/world-us-canada-42254219

Bennett, W. L. (2012). The personalisation of politics: Political identity, social media, and changing patterns of participation. *The Annals of the American Academy of Political and Social Science*, *644*, pp. 20–39.

Brooks, D. (2019). The cruelty of call-out culture. The New York Times. Retrieved from https://www.nytimes.com/2019/01/14/opinion/call-out-social-justice.html

Brown, R. (1975). *Strain of violence*. New York, NY: Oxford University Press.

Burrows, W. (1976). *Vigilante*. NY: Harcourt Brace Jovanovich.

Castells, M. (2008). Interview with Manuel Castells. *Chinese Journal of Communication*, *1*, pp. 3–6.

Castells, M. (2011). *Communication power*. Oxford: Oxford University Press, pp. 3–348.

Chang, L. Y. C., & Poon, R. (2017). Internet vigilantism: Attitudes and experiences of university students toward cyber crowdsourcing in Hong Kong. *International Journal of Offender Therapy and Comparative Criminology*, *61*(16), pp. 1,912–1,932. https://doi.org/10.1177/0306624X16639037

Channel News Asia. (2019). As the 'human flesh search engine' of online vigilantism grows, measures to deal with doxxing critical, MPs say. Channel News Asia. Retrieved from https://www.channelnewsasia.com/news/singapore/human-flesh-search-engine-of-online-vigilantism-doxxing-poha-11511328

Cheong, P. H., & Gong, J. (2010) Cyber vigilantism, transmedia collective intelligence, and civic participation. *Chinese Journal of Communication*, *3*(4), pp. 471–487.

Cheow, S. (2019). Victim of 'minor intrusion' molestation case disappointed at court's decision. The New Paper. Retrieved from https://www.tnp.sg/news/singapore/victim-molestation-case-disappointed-courts-decision?utm_medium=Social&utm_source=Facebook&fbclid=IwAR2KnoAaibvIOBdlfm4yWZkNg1NDruIKVQWLo0Fr76PT7WfdVmDm10ySj4#Echobox=1569542986

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior*, *24*(6), pp. 659–683. https://doi.org/10.1023/A:1005552203727

Douglas, D. M. (2016). Doxing: A conceptual analysis. *Ethics and Information Technology*, *18*(3), pp. 199–210.

Dutta, M. J. (2019). Sexual harassment is the norm at university. Singapore student Monica Baey sent us a wake-up call. *South China Morning Post*. Retrieved from https://www.scmp.com/week-asia/opinion/article/3009357/sexual-harassment-norm-university-singapore-student-monica-baey

Fenton, N. (2016) Left out? Digital media, radical politics and social change. *Information, Communication & Society*, *19*(3), pp. 346–361.

Fu, W. W., & Sim, C. C. (2011). Aggregate bandwagon effect on online videos' viewership: Value uncertainty, popularity cues, and heuristics. *Journal of the American Society for Information Science and Technology*, *62*(12), pp. 2382–2395. https://doi.org/10.1002/asi.21641

Ganapathy, N. (2017). Rehabilitation, reintegration and recidivism: A theoretical and methodological reflection. *Asia Pacific Journal of Social Work and Development*, *28*(3), pp. 153–167.

Henderson, R. (2019). 5 Reasons why people love cancel culture. Psychology Today. Retrieved from https://www.psychologytoday.com/sg/blog/after-service/201912/5-reasons-why-people-love-cancel-culture

Jagdish, B. (2018). Is online vigilantism the best way to exact social justice? Channel News Asia. Retrieved from https://www.channelnewsasia.com/news/commentary/caltex-bmw-driver-petrol-pump-attendant-vigilantism-social-justi-10150096

Johnston, L. (1996). What is Vigilantism? *The British Journal of Criminology*, *36*(2), pp. 220–236. https://doi.org/10.1093/oxfordjournals.bjc.a014083

Jubbal V. (2016). Experience: I was accused of carrying out the Paris attacks. Retrieved from https://www.theguardian.com/lifeandstyle/2016/jul/01/experience-i-was-accused-of-carrying-out-the-paris-attacks

Khair, M. (2018). Three Factors that have kept Recidivism Rates Low in 2017. Home Team News. Retrieved from https://www.mha.gov.sg/hometeamnews/our-community/ViewArticle/three-factors-that-have-kept-recidivism-rates-low-in-2017

Koh, V. (2016). More flexibility in community-based sentencing possible: Shanmugam. Today. Retrieved from https://www.todayonline.com/singapore/more-flexibility-community-based-sentencing-possible-shanmugam

Langone, A. (2018). #MeToo and Time's Up Founders Explain the Difference Between the 2 Movements—And How They're Alike. Time. Retrieved from http://time.com/5189945/whats-the-difference-between-the-metoo-and-times-up-movements/

Lim, K. (2019). How young women are using social media to fight back against men behaving badly. Today. Retrieved from https://www.todayonline.com/singapore/how-young-women-are-using-social-media-fight-back-against-men-behaving-badly?cid=h3_referral_inarticlelinks_03092019_todayonline

Lin, M. (2017). Online 'CSI' vigilantes: The good, the bad and the ugly. The Straits Times. Retrieved from https://www.straitstimes.com/singapore/online-csi-vigilantes-the-good-the-bad-and-the-ugly

McCann, K. (2011). Vancouver Police Department 2011 Stanley Cup Riot Review, 101 MeTooMvmt.org (n.d.). Retrieved from https://metoomvmt.org/

Milieu (2019). Most Singaporeans Support the Criminalisation of Doxxing. Retrieved from https://mili.eu/insights/most-singaporeans-support-the-criminalisation-of-doxxing

Ministry of Law (2019, April 1). Enhancements to the Protection from Harassment Act [Press release]. Retrieved from https://app.mlaw.gov.sg/news/press releases/enhancements-to-the-protection-from-harassment-act-poha

Ministry of Law (2019b, December 27). Commencement of Amendments to the Penal Code and Other Legislation on 1 January 2020 [Press release]. Retrieved from https://www.mlaw.gov.sg/news/press-releases/commencement-of-amendments-to-the-penal-code-and-other-legislation-on-1-january-2020

Mokhtar, F. (2017). One-stop centre among measures to lessen trauma of sex-crime victims. Today. Retrieved from https://www.todayonline.com/singapore/one-stop-centre-among-new-measures-reduce-stress-trauma-sex-crime-victims-mha

Nadeau, R., Cloutier, E., & Guay, J.-H. (1993). New Evidence About the Existence of a Bandwagon Effect in the Opinion Formation Process. *International Political Science Review*, *14*(2), pp. 203–213. https://doi.org/10.1177/019251219301400204

Neo R.W. (2019). A look at key changes to Protection from Harassment Act. Today. Retrieved from https://www.todayonline.com/singapore/look-key-amendments-protection-harassment-bill

Ng, H. (2019). NUS Peeping Tom case: Monica Baey urges online bullying against Nicholas Lim to stop. The Straits Times. Retrieved from https://www.straitstimes.com/singapore/nus-peeping-tom-case-monica-baey-urges-online-bullying-against-nicholas-lim-to-stop

Ng, H., & Choo, Y. T. (2019). NUS to convene review committee after student calls for tougher action against man who filmed her in shower. The Straits Times. Retrieved from https://www.straitstimes.com/singapore/nus-to-convene-review-committee-after-student-calls-for-tougher-action-against-man-who

Nhan, J., Huey, L., & Broll, R. (2017). Digilantism: An analysis of crowdsourcing and the Boston marathon bombing. *The British Journal of Criminology*, *57*, pp. 341–361.

Online Hate Prevention Institute (2015). Online Vigilantism and its perils. Online Hate Prevention Institute. Retrieved from https://ohpi.org.au/online-vigilantism-and-its-perils/

Pan, J. (2020). 16 Questions with The Founder of "SG Covidiots." Rice Media. Retrieved from https://www.ricemedia.co/current-affairs-features-16-questions-sg-covidiots/

Rappaport, J. (1981). In praise of paradox: a social policy of empowerment over prevention. *American Journal of Community Psychology*, *9*, pp. 1–25.

Robbins, S. P. (2015). From the Editor—The Red Pill or the Blue Pill? Transcending Binary Thinking, *Journal of Social Work Education*, *51*(1), pp. 1–4.

Romano, A. (2019). Why we can't stop fighting about cancel culture. Vox. Retrieved from https://www.vox.com/culture/2019/12/30/20879720/what-is-cancel-culture-explained-history-debate

Silva, K. K. E. (2017). Vigilantism and cooperative criminal justice: Is there a place for cybersecurity vigilantes in cybercrime fighting? *International Review of Law, Computers & Technology*, *43*(1), pp. 21–36.

Sim, D. (2019). Singaporeans decry judge's move to spare NUS student Terence Siow from jail for molest charge. South China Morning Post. Retrieved from https://www.scmp.com/week-asia/politics/article/3030712/singaporeans-decry-judges-decision-spare-nus-student-jail-term

Soon, C. & Cho, H. (2014). OMGs! Offline-based movement organisations, online-based movement organisations and network mobilisation: A case study of political bloggers in Singapore. *Information, Communications & Society*, *17*(5), pp. 537–559.

Tang, L. (2020). Online shaming of those flouting Covid-19 circuit breaker rules could amount to doxxing, say lawyers. Today. Retrieved from https://www.todayonline.com/singapore/online-shaming-those-flouting-covid-19-circuit-breaker-rules-could-amount-doxxing-say

Tam, L. (2018). Why stalking, cyberbullying and doxxing are so harmful and what makes people do it. South China Morning Post. Retrieved from https://www.scmp.com/lifestyle/family-relationships/article/2165543/why-stalking-cyberbullying-and-doxing-are-so-harmful

Tay, X. Y. (2019). Don't punish victims for naming abusers. The Straits Times. Retrieved from https://www.straitstimes.com/forum/letters-on-the-web/weeks-top-letter-2-dont-punish-victims-for-naming-abusers

Teng, A. (2019a). NUS Peeping Tom given conditional warning due to high likelihood of rehabilitation: Police. The Straits Times. Retrieved from https://www.straitstimes.com/singapore/courts-crime/student-in-nus-sexual-misconduct-case-given-conditional-warning-due-to-high

Teng, A. (2019b). Local universities reviewing how they handle sexual misconduct and support for victims. The Straits Times. Retrieved from https://www.straitstimes.com/singapore/education/smu-reviewing-disciplinary-processes-for-sexual-misconduct-says-every-complaint

Thian, Y. S. (2014). The four principles that anchor Singapore's criminal justice system. Law Gazette. Retrieved from https://v1.lawgazette.com.sg/2014-02/964.htm

Thirumaran, K. J. (2019). The evolution of the Singapore criminal justice process. *Singapore Academy of Law Journal*, pp. 1,042–1,067.

Tomer-Fishman, T. (2010). Cultural Defense, Cultural Offense, or No Culture at All: An Empirical Examination of Israeli Judicial Decisions in Cultural Conflict Criminal Cases and of the Factors Affecting Them. *Journal of Criminal Law and Criminology*, *100*(2), pp. 475–521.

Van Ness, D. W., & Strong, K. H. (2006). *Restoring Justice: An Introduction to Restorative Justice* (3rd ed.), Newark, NJ: LexisNexis Matthew Bender.

Vijayan, K. C. (2018). Parliament: Proposed criminal justice reforms a 'major step', says MinLaw. The Straits Times. Retrieved from https://www.straitstimes.com/politics/parliament-proposed-criminal-justice-reforms-a-major-step-towards-a-system-that-still

Wong, L. (2019). Laws must keep pace with tech trends to safeguard citizens' privacy and rights, experts say. The Straits Times. Retrieved from https://www.straitstimes.com/tech/laws-must-keep-pace-with-tech-trends-to-safeguard-citizens-privacy-and-rights-experts-say

Wong, Y., & Loh, L. S. P. (2019). Understanding the Phenomenon of Cyber Vigilantism (HTBSC Research Report 17/2019). Singapore: Home Team Behavioural Sciences Centre.

Zetter, K. (2007). Cyberbullying suicide stokes the Internet fury machine. Wire. Retrieved from http://archive.wired.com/politics/onlinerights/news/2007/11/vigilantejustice

Zheng, Z. (2019). NUS student Monica Baey returns to Taiwan & writes closure post on Instagram. Mothership. Retrieved from https://mothership.sg/2019/05/monica-baey-closure-ig-post/

Zimmerman, M. (1995). Psychological empowerment: Issues and illustrations. *American Journal of Community Psychology*, *23*, pp. 581–599.

Zimmerman, M. (2000). Empowerment Theory. Psychological, Organizational and Community Levels of Analysis. In J. Rappaport & E. Seidman (Eds.), Handbook of Community Psychology. New York: Kluwer Academic, pp. 43–63.

Zimmerman, M., Israel, B., Schulz, A., & Checkoway, B. (1992). Further explorations in empowerment theory: An empirical analysis of psychological empowerment. *American Journal of Community Psychology*, *20*, pp. 707–727.

https://doi.org/10.1142/9789811232411_0008

# Chapter 8

# Understanding the Growing Prevalence of Information Operations on Social Media

Xingyu Ken Chen* and Jessie Janny Thenarianto†

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

**Xingyu_CHEN_from.TP@mha.gov.sg*

†*Jessie_THENARIANTO@mha.gov.sg*

## 8.1 Introduction

At the end of the 2010s, it became clear that there were growing numbers of organised information operations aimed at manipulating online discourse on social media [Bradshaw & Howard, 2019]. Information operations (IO) refers to the purposeful and systematic generation and dissemination of information by governments or non-state actors to distort domestic or foreign political sentiments in order to achieve a strategic and/or geopolitical outcome [Weedon *et al*., 2017]. Fundamentally, IO activities are hostile to the national interest of the targeted country—as they are often used to interfere with elections, sow social discord, and weaken political support for policies perceived as adversarial for the IO actor [see Mueller, 2019; Pomerantsev & Weiss, 2014].

From a criminological perspective, the conduct of IO can be seen as a kind of cyber-deviance that violates the norms and beliefs of a larger

culture [Holt & Bossler, 2016], where the technological affordances of social media are being misused to undermine the quality of public conversations online [Roth, 2019]. Furthermore, with the rapid advancements in information communication technologies such as social media, even IO actors with few resources can dramatically scale up their capability to influence their target audience. By exploiting the affordances of social media, these IO campaigns can be executed at a low cost and with greater plausible deniability for its creators [see Beskow & Carley, 2019; Diamond & Schell, 2019; Green, 2016; Weedon *et al*., 2017].

According to a 2019 Oxford Internet Institute study, the researchers found evidence of IO taking place in 70 countries—on every single continent except for Antarctica [Bradshaw & Howard, 2019]. Even Southeast Asia as a region is not immune to the threat of IO—there was evidence of IO being carried out in a majority of Southeast Asian countries such as Indonesia, Malaysia, the Philippines, Thailand, Cambodia, Myanmar, and Vietnam [Abdul Rahman *et al*., 2020; Bradshaw & Howard, 2019].

One of the most prominent IO campaigns in recent memory was the 2016 US elections, where a Russia-linked company known as the Internet Research Agency (IRA) was linked to attempts to interfere with the US elections through efforts such as spreading messages on social media aimed at harming the Clinton campaign [Mueller, 2019]. In Hong Kong, there was evidence of a coordinated effort on Twitter to delegitimise the Hong Kong protest movement which broke out due to opposition to the extradition bill ["Information operations directed at Hong Kong," 2019; Uren *et al*., 2019]. In Indonesia, police announced that they arrested key members of a fake news syndicate known as Saracen, who was linked to online falsehoods concerning the former Jakarta governor Basuki 'Ahok' Tjahaja Purnama as well as President Joko Widodo and his allies [Chan, 2017; Maulia & Nugroho, 2017]. These falsehoods were often of a racial or religious nature, which could easily stoke ethnic tensions in Indonesia.

The growing prevalence of IO around the world illustrates a threat posed by IO actors who can and have demonstrated their capabilities in influencing public opinion using the internet. These actors can provoke and amplify social tensions, as well as muddle public decision-making in societies. Furthermore, Chen and Chin [2019] argue that these actors appear to possess a working knowledge of marketing, psychology, mass communications, and existing political issues which is utilised to maximise the effects of their operations.

Therefore, these factors and considerations present a need to understand why these IO are so compelling from a psychological

perspective. As such, this chapter seeks to understand how IO actors succeed in manipulating public opinions online during crucial times, such as during elections. This psychological understanding is the first step towards constructing evidence-centric policies to mitigate the threat of IO.

### 8.1.1 *Conditions that enabled IO to flourish*

While the conduct of IO is not new, recent conditions have enabled the increasing prevalence of IO. Some of these factors are: (1) the recent successes of IO emboldening other actors; (2) existing social fault lines and identity politics practices; and (3) the conduct of IO-enabling actors with different motivations to fulfil their goals/aims.

### 8.1.2 *The recent successes of IO emboldening other actors*

The success of IO in recent political events around the world, such as the 2016 US presidential elections [Allcott & Gentzkow, 2017] and the 2013 Malaysian general elections [Guest, 2018], demonstrated that these events are vulnerable to IO activity. This is consistent with the opportunity perspective in criminology—stating that offenders make decisions related to crime events based on what they perceive as an opportunity [Wilcox *et al*., 2012]. Moreover, the social learning theory in psychology states that individuals learn from observing behaviour, and that behaviours which produce valued rewards attract individuals the most [Bandura & Walters, 1977]. Arguably, highly publicised IO successes therefore can act as a template for other perpetrators to adopt and learn from.

In the lead up to the 2016 US presidential elections, the IRA used "coordinated disinformation tactics" to influence US citizens' vote choice and intensify social divisions in the country [DiResta *et al*., 2018]. The company used popular social media platforms (e.g., Facebook, Twitter, Instagram, YouTube) to spread pro-Trump and anti-Hillary Clinton content, as well as attempting to exacerbate existing fault lines such as immigration and LGBT issues, and gun rights [Howard *et al*., 2018]. Their posts had a wide reach, generating 77 million user engagements on Facebook, 187 million user engagements on Instagram, and 73 million user engagements on Twitter [DiResta *et al*., 2018]. Researchers believe that a large number of posts have been distributed, and continue to be

circulated in the online sphere [DiResta *et al*., 2018]—all of which can be viewed as an indicator of success.

Moreover, the successful execution of IO campaigns for prior elections could prompt more entrepreneurial IO actors to monetise their capability to conduct IO ["Kaiser expose," 2020; Paramaesti, 2018]. Recent investigations into specific organisations that carry out IO for-profit found several examples. IO can be a feasible business model as there is demand from certain parties that are willing to pay for such services. For example, Saracen, the fake news syndicate based in Indonesia, allegedly sent a social media campaign proposal containing website-making and buzzer services to a prospective buyer [Paramaesti, 2018]. Likewise, in the Philippines, research had uncovered a network of operators who formulate disinformation campaigns and mass cyber armies for interested political actors [Gleicher, 2019b; Ong & Cabañes, 2018].

### 8.1.3 *Existing social fault lines and identity politics practices*

Another factor that has enabled these IO actors to thrive is the existing societal fault lines and identity politics in society, which they were able to exploit. Some common fault lines used by these actors are often related to social identities, such as religion, race, and social class [Arif *et al*., 2018].

Divisive issues involving religion, race, nationality, and gender can easily trigger emotions among the in-group and increase the likelihood of conflicts [Chrobot-Mason *et al*., 2009; Stephan *et al*., 2009]. Exploiting issues related to social identities is an effective strategy for IO actors since these issues are emotionally laden and IO actors can easily play up the threat posed by the out-group.

For instance, Muhammad Faizal Tonong, a member of Saracen, tapped onto existing religious and socially-divisive issues by uploading a meme saying that Indonesian President Joko Widodo "won't sell [in] 2019, [he is] clearly anti-Islam pro-PKI,[a] and doesn't keep a lot of his

[a]PKI stands for Partai Komunis Indonesia, a communist political party in Indonesia that was banned following its alleged involvement in a coup attempt in 1965 (Boden, 2007). Ever since the 1965 attempt or the 30 September November, some have observed that Indonesia has developed an "anti-communist" paranoia (Lamb, 2017). Moreover, according to these observers labelling individuals as being a communist has been the classic "go-to method of attack" for decades in Indonesia (Azali, 2017).

promises" [Andriyanto, 2017, para. 4]. Another meme posted by him mentioned that the Indonesian National Police Chief was behind the conflict between Banser Nahdlatul Ulama[b] and Hizbut Tahrir Indonesia[c] [Andriyanto, 2017]. Both posts aimed to undermine public trust in authority figures by inciting outrage based on religious issues.

Similarly, the Malaysian cybertrooper, Syarul Ema Rena and her colleagues, amplified racial and religious sensitivities through their work in creating fake accounts, impersonating supporters of the opposition, and posting racist statements [Guest, 2018]. During the 2014 Teluk Intan by-election, she fabricated a video of an Indian member of the Barisan Nasional (BN) coalition, whom in the video claimed that an activist from the opposition had attacked him. The activist was alleged to have also made offensive remarks about the BN's Indian member's mother and religion [Leong, 2019]. The video, which aimed to sway the Indian minority voters, made rounds on WhatsApp 24 hours before the voting started [Leong, 2019].

In the US, it was found that the IRA polarised public discussion surrounding the Black Lives Matter movement. This social movement was concerned about issues such as the systemic racism and violence experienced by African-Americans [Arif *et al*., 2018]. The IRA infiltrated both sides of the movement and incited notions of violent retribution against the other side [Arif *et al*., 2018; Devine, 2017].

These examples show how IO campaigns amplify the fear and outrage in the public by highlighting the need to defend the in-group (e.g., being a member of a particular religion) from the threat posed by the target. Furthermore, pre-existing practices of politicising identity in society make it easier for certain IO narratives to take root. The politicisation of identity, which is sometimes referred to as identity politics, is the political mechanism of using identity as a source and tool for politics [Haboddin, 2012].

In Indonesia, researchers found that political strategies emphasising identity can be traced back to the 1990s, when several provinces insisted on prioritising their own natives in governmental positions [Haboddin, 2012]. The use of religious issues for political marketing strategies

[b] Banser Nahdlatul Ulama is an autonomous body affiliated with Nahdlatul Ulama, one of the largest Islamic organisations in Indonesia ("Sejarah banser," 2018).

[c] Hizbut Tahrir Indonesia (HTI) is the Indonesian chapter of the larger Hizbut Tahrir organisation, which aims to restore the Islamic caliphate (Osman, 2010). It is was disbanded by the Indonesian government in 2017 (Sinaga *et al*., 2017).

(e.g., calling for voters to choose leaders who share the same religious background as them) have also been found in many local elections during 2015 to 2018 [Ramadlan & Masykuri, 2018]. Similarly, Malaysian politics have also been characterised by "racially charged rhetoric" since its independence from the UK [Sukumaran, 2017]. Even for the US, racial identity politics have been exploited by foreign actors to incite racial hatred in the past. During the Cold War period, Soviet intelligence fabricated conspiracies that the assassination of Martin Luther King Jr., a prominent African-American civil rights activist, was the result of a plan between white racists and the US government [Andrew & Mitrokhin, 2000]. As such, existing identity politics in any society creates a vulnerability that IO actors can exploit by fanning these social divisions.

## 8.2 The Conduct of IO Enabling Actors with Different Motivations to Fulfil their Goals/Aims

There are various reasons why people carry out IO, and the conduct of IO enables actors with very different motivations to meet their goals or aims. Three types of motivations are identified here: (1) political gain; (2) financial gain; and (3) a mixture of multiple motives.

Prior studies have identified political or ideological motivation as a factor influencing IO [Lewis & Marwick, 2017]. These actors engage in IO out of their ideological commitment to a cause or due to a desire to enable a particular political party to take power. For instance, during the 2016 US presidential elections, Ovidiu Drobota, the Romanian man behind 'Ending The Fed', a website which was behind four[d] of the top 10 false election stories on Facebook, claimed that his support towards Trump was what prompted him to launch the site [Townsend, 2016]. In Malaysia, Syarul Ema Rena asserted that she was never paid. She wrote and amplified political falsehoods since she was ideologically committed to the ruling coalition [Guest, 2018].

Financial gain is another motive for IO actors. For instance, the Macedonian students producing stories in favour of Trump were purely

[d]The four fake stories were: Pope Francis endorsing Donald Trump, Hilary Clinton selling weapons to ISIS, Hillary Clinton being disqualified from holding federal office, and the FBI director receiving millions from the Clinton Foundation.

driven by monetary motivations, profiting from advertising revenues derived from creating websites circulating such stories [Allcott & Gentzkow, 2017; Morgan, 2018]. Paul Horner, the lead writer behind a fake news site, National Report, reportedly wrote pro-Trump stories for financial gain while publicly declaring that he opposed Trump [Allcott & Gentzkow, 2017].

Some IO actors can also be motivated by a mix of political and financial motivations. For example, Saracen, the syndicate that organised a disinformation campaign containing fake news articles about Ahok during his candidacy as governor, ran a business of hoax creation and propagation for profit [Chandran, 2017]. For example, in a separate event, Saracen had allegedly sent a proposal to prospective clients running for office with fees of up to 72 million rupiahs [Arela *et al*., 2017]. Separately, Saracen's leader, Jasriadi, was a known supporter of Prabowo Subianto, Joko Widodo's opponent in the 2014 and 2019 presidential elections [Persada *et al*., 2017]. Jasriadi also admitted that he had hacked accounts belonging to those who were attacking Islam and Prabowo [Persada *et al*., 2017].

Ong and Cabañes [2018] found that groups behind disinformation campaigns in the Philippines consisted of elite advertising and public relations (PR) strategists, digital influencers, key opinion leaders, and, community-level fake account operators. While anonymous digital influencers and community-level fake account operators tend to be financially motivated, advertising and PR strategists viewed such operations as a new challenge for them to apply their skills and leverage their networks [Ong & Cabañes, 2018].

## 8.3 Social Media Strategies, Tools, and Techniques Used by IO Actors

While there is nothing new about IO, their tactics have evolved with the times. As social media platforms such as Twitter and Facebook become commonly used spaces for the public to discuss politics, disseminate news, and organise collective action [An *et al*., 2014; Arif *et al*., 2018], this development makes it an attractive platform for IO actors to carry out their campaigns.

As such, there is a recent shift towards IO actors operating on social media as the affordances of the technology enable them to enhance the scale and precision of the propaganda they wish to spread [Bradshaw & Howard, 2019]. Furthermore, IO actors have developed tactics that are

unique to social media. They can impersonate genuine grassroots voices by using fake accounts as a front as well as amassing armies of fake accounts to amplify their messages artificially [Arif *et al*., 2018; Chen & Chin, 2019].

### 8.3.1 *Impersonation tactics: Masking identities with fake accounts*

IO actors often mask their true identities through fake accounts. In particular, the ease of setting up new accounts on social media under a variety of false aliases makes this strategy particularly attractive for IO actors. The use of fake accounts can provide some degree of plausible deniability, as well as making it difficult for authorities to detect the actual persons behind the campaign [Abdul Rahman *et al*., 2020; Green, 2016; Kragh & Åsberg, 2017].

Moreover, these accounts can be used to impersonate authentic grassroots voices to amplify more support for a political candidate [Keller *et al*., 2020] or drown out opposition voices [King *et al*., 2017]. For example, it was found that the IRA was operating right-wing troll accounts as well as left-wing troll accounts to influence various U.S. political events such as the Black Lives Matter movement, spreading alt-right narratives, and the U.S. elections [Linvill & Warren, 2020]. Some have noted that these activities were aimed at sowing political discord in the U.S. [Howard *et al*., 2018; Linvill & Warren, 2020; Mueller, 2019].

The sophistication and variety of these methods can vary depending on the groups. Some may opt to operate fan pages or closed social media groups to attract an audience and then surreptitiously spread propaganda messages to them. In the Philippines, it was reported that a Korean-Pop fan page for South Korean actor Kim Soo-hyun was being hijacked to spread political propaganda [Occeñola, 2018]. Administrators of conspiracy theories groups such as the Filipino Flat Earth group were also known to have weaved both conspiracy theories as well as political propaganda when posting to the Filipino Flat Earth group [Ong *et al*., 2019].

In other cases, some IO actors may hack social media accounts in order to repurpose these accounts for propaganda purposes. They could also hack the social media accounts of their opponents to undermine the opponent's outreach efforts [Bradshaw & Howard, 2019]. This was seen in the case of Saracen, where Jasriadi was arrested for illegally hacking into social media accounts and accused of using them to spread fake news

and hate speech [Harahap, 2018]. Jasriadi admitted to hacking accounts on the basis that these accounts were attacking Islam and Prabowo [Persada *et al*., 2017].

Impersonation tactics can help increase the persuasiveness of IO messages. IO actors can be more persuasive when they send their messages using multiple accounts posing as different sources of information. Research has found that multiple sources are more convincing than a single source, especially if those sources contain varying arguments that point to the same conclusion [Paul & Matthews, 2016].

### 8.3.2 *Amassing accounts for the conduct of IO*

Unlike traditional media, social media has specific affordances that enable IO actors to change the scale, scope, and precision of their information campaigns in a cost-effective manner. One example of this is how IO actors can amass and operate large numbers of fake accounts with just a small outfit. In Malaysia, it was estimated that small groups of cyber troopers were operating thousands of fake accounts [Leong, 2019]. In the case of Saracen, five people have been arrested by Indonesian authorities in connection to Saracen [Chan, 2017] and 800 Saracen-linked Facebook accounts, 207 pages, 546 groups, and 208 Instagram accounts were subsequently shut down by Facebook [Gleicher, 2019a].

Moreover, IO actors often react to political events by quickly amassing accounts or bots. As a result, the social media accounts spreading IO content are relatively young as they are often created for specific campaigns in mind. For example, Grimme *et al.* [2018] found that accounts attempting to influence perceptions of a German elections television debate were less than one month old. Similarly, in the case of the accounts trying to delegitimise the 2019 Hong Kong protests, it was found that most of these accounts were created within 2019 [Chua, 2019; Uren *et al*., 2019].

Due to the need to effectively reach out to a wider audience, many accounts will have to be simultaneously created and/or reassigned at the beginning of a campaign [Keller *et al*., 2020]. In some cases, IO actors may even resort to the use of bots to artificially amplify the reach of such campaigns in an automated fashion [see Ferrara, 2017; Varol *et al*., 2017].

From a psychological perspective, this tactic of amassing accounts is aimed at enabling IO actors to create the illusion of widespread support or criticism for a person, organisation or cause [Bienkov, 2012].

This method, also known as astroturfing, is useful for IO actors as it enables them to amplify the reach of their narratives artificially. Astroturfing leverages on people's tendencies to look for 'popularity cues' to what they read online [Fu & Sim, 2011], as well as to conform to popular opinions held by their in-group [Colleoni *et al*., 2014].

## 8.4 Implications

Broadly speaking, understanding how IO actors mount campaigns to manipulate online opinion during elections highlights the need for approaches to limit the effectiveness of IO actors. Two approaches to limit the effectiveness of IO actors are discussed here. The first approach is to enable the public to resist influence efforts by IO actors. The other approach is for authorities to enact transparency measures to make it difficult for IO actors to operate without subterfuge.

### 8.4.1 *Build capacity in the public to resist IO*

Current evidence indicates that countries who invested in media literacy were well equipped to resist disinformation and their ramifications [Charlton, 2019; Lessenski, 2018]. As an example, Finland is widely lauded for their ability to resist the wave of propaganda from Russia, which has been attributed to the country's long history of dealing with Russian influence, its strong public education system, and effective government strategy to deal with disinformation [Standish, 2017].

In Singapore, there are ongoing literacy efforts, such as the S.U.R.E campaign that seeks to increase information literacy in the public by educating the public to check news source and to assess it for its accuracy [Tan *et al*., 2014]. This approach can be extended in various ways to increase the public's ability to resist an influence effort by IO actors. Given that IO actors are constantly devising a myriad of techniques to disguise their identity, it might not be realistic to educate the public about all the techniques used. However, if people are encouraged to be sceptical and to check the source behind the information, it may be possible for them to detect and reject such influence efforts [Wineburg & McGrew, 2017].

Another way to build up the public's capacity to resist IO efforts is by strengthening intergroup relations within the society. Given that social fault lines and identity politics are exploited by IO actors to sow social discord, efforts to bridge such fault lines in society can act as a protective

factor to buffer the effects of an IO. In multi-ethnic and multi-cultural societies such as Singapore, there are ongoing efforts to form networks of trust between people of different races and religious groups [Latif, 2011], which can be an effective means of undermining the efforts of IO actors seeking to exploit such fault lines. In Singapore, one avenue of building such networks is the Inter-Racial and Religious Confidence Circles (IRCC), which are local-level inter-faith platforms formed to promote racial and religious harmony in a melting pot society such as Singapore's ["Inter-Racial and Religious Confidence Circles (IRCCs)," 2019].

### 8.4.2 *Transparency measures to hinder IO actors*

As a principle, transparency has been described as 'sunlight' which disinfects a society against manipulative activity [Diamond & Schell, 2019; "Speech introducing the National," 2017]. Measures that expose the actions or the people trying to manipulate public opinion can undermine the efforts of these malicious actors [Diamond & Schell, 2019]. Given that IO actors benefit from tactics that grant them some degree of plausible deniability, transparency measures could make it difficult for them to operate and possibly deter them from trying.

For example, legislative measures can target and unmask those who are carrying out IO. Australia has launched a parliamentary inquiry to examine how hostile actors can use social media to interfere with its elections [Packham, 2019]. In a similar vein, Singapore has highlighted the need for legislation to combat foreign interference, citing the importance of early detection and exposure of such activities [Koh, 2019]. As of the writing of this chapter, the aforementioned legislation has not been passed.

Tech companies such as Google, Facebook, and Twitter have implemented transparency measures on their platforms to discourage IO actors from operating on their platforms. These measures can range from increasing public understanding how IO is being carried out [Harrison, 2019], showing which parties are paying for political advertisements [Stefano, 2019], as well as documenting the various ways which IO actors conceal their influence efforts after they were shut down [Perkins, 2019]. For example, Twitter has taken to naming and shaming IO actors by releasing datasets of deleted accounts linked to IO for public scrutiny ["Information Operations Directed at," 2019; Roth, 2019]. Facebook and Google have granted access to their databases, which allows scrutiny to the kinds of political advertising done on their platforms ["Facebook, Google Tools Reveal," 2019].

These efforts raise the cost of conducting IO as influence efforts can backfire when the perpetrators' actions are being exposed. Consequently, it is clear that social media companies are essential partners in identifying and exposing IO activity on their platforms. Additionally, such measures can help to prevent IO actors from profiting from IO [Love & Cooke, 2016], or by shutting down these actors before they can inflict much damage ["Information operations directed at Hong Kong," 2019]. Furthermore, it highlights the importance of understanding how these IO actors are operating deceptively so that such an understanding can inform policies seeking to curb such behaviour.

## 8.5 Acknowledgement

The views expressed in this chapter are those of the authors only and do not represent the official position or view of the Ministry of Home Affairs (MHA), Singapore.

## 8.6 References

Abdul Rahman, M. F., Haciyakupoglu, G., Ang, B., Leong, D., Yang, J., & Teo, Y.-L. (2020). *Cases of foreign interference in Asia*. Centre of Excellence for National Security.

Allcott, H., & Gentzkow, M. (2017). *Social media and fake news in the 2016 election* (NBER Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w23089

An, J., Quercia, D., Cha, M., Gummadi, K., & Crowcroft, J. (2014). Sharing political news: The balancing act of intimacy and socialization in selective exposure. *EPJ Data Science*, *3*(1), p. 12.

Andrew, C., & Mitrokhin, V. (2000). *The sword and the shield: The Mitrokhin archive and the secret history of the KGB*. Hachette UK.

Andriyanto, H. (2017, July 21). Ini isi akun Faizal Tonong yang membuatnya ditangkap. *BeritaSatu*. https://www.beritasatu.com/nasional/442894/ini-isi-akun-faizal-tonong-yang-membuatnya-ditangkap

Arela, G., Durohman, I., & Yurisa, R. G. (2017, August 28). Penyalur guru nyambi penyalur kebencian. *Detik.Com*. https://news.detik.com/x/detail/investigasi/20170828/Penyalur-Guru-Nyambi-Penyalur-Kebencian/

Arif, A., Stewart, L. G., & Starbird, K. (2018). Acting the part: Examining information operations within #BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), p. 20.

Azali, K. (2017). Fake news and increased persecution in Indonesia. *ISEAS Perspective*, *2017*(61), 1–10. https://www.iseas.edu.sg/images/pdf/ISEAS_Perspective_2017_61.pdf

Bandura, A., & Walters, R. H. (1977). *Social learning theory* (Vol. 1). Prentice-hall Englewood Cliffs, NJ.

Beskow, L. C. D. M., & Carley, K. M. (2019). Social cybersecurity: An emerging national security requirement. *Military Review*, *99*(2), p. 117.

Bienkov, A. (2012, February 8). Astroturfing: What is it and why does it matter? *The Guardian*. https://www.theguardian.com/commentisfree/2012/feb/08/what-is-astroturfing

Boden, R. (2007). The 'Gestapu' events of 1965 in Indonesia: New evidence from Russian and German archives. Bijdragen tot de Taal-, Land- en Volkenkunde (BKI), *163*(4), 507–528.

Bradshaw, S., & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Project on Computational Propaganda.

Chan, F. (2017, September 17). Indonesian police uncover "fake news factory." *The Straits Times*. https://www.straitstimes.com/asia/se-asia/indonesian-police-uncover-fake-news-factory

Chandran, N. (2017, December 26). Fake news was a weapon in Asia in 2017. *CNBC*. https://www.cnbc.com/2017/12/25/fake-news-was-a-weapon-in-asia-in-2017.html

Charlton, E. (2019). How Finland is fighting fake news in the classroom. *World Economic Forum.* https://www.weforum.org/agenda/2019/05/how-finland-is-fighting-fake-news-in-the-classroom/

Chen, X. K., & Chin, J. (2019). *A behavioural sciences perspective on common tactics used by hostile information campaigns actors during the Hong Kong protests* (HTBSC Research Report S11/2019). Home Team Behavioural Sciences Centre.

Chrobot-Mason, D., Ruderman, M. N., Weber, T. J., & Ernst, C. (2009). The challenge of leading on unstable ground: Triggers that activate social identity faultlines. *Human Relations*, *62*(11), pp. 1,763–1,794.

Chua, C. H. (2019, August 20). Target HK: A quick dive into China's disinformation campaign on Twitter. *Medium.* https://towardsdatascience.com/target-hk-a-quick-dive-into-chinas-disinformation-campaign-on-twitter-2b64ab9feb1a

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), pp. 317–332.

Devine, C. (2017, October 31). 'Kill them all'—Russian-linked Facebook accounts called for violence. *CNN*. https://money.cnn.com/2017/10/31/media/russia-facebook-violence/index.html

Diamond, L., & Schell, O. (2019). *China's influence and American interests: Promoting constructive vigilance*. Hoover Press.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J., & Johnson, B. (2018). *The tactics & tropes of the Internet Research Agency*. New Knowledge.

Facebook, Google tools reveal new political ad tactics. (2019, March 5). *CNA*. https://www.channelnewsasia.com/news/world/facebook--google-tools-reveal-new-political-ad-tactics-10904726

Ferrara, E. (2017). *Disinformation and social bot operations in the run up to the 2017 French presidential election* (SSRN Scholarly Paper ID 2995809). Social Science Research Network. https://papers.ssrn.com/abstract=2995809

Fu, W. W., & Sim, C. C. (2011). Aggregate bandwagon effect on online videos' viewership: Value uncertainty, popularity cues, and heuristics. *Journal of the American Society for Information Science and Technology*, *62*(12), pp. 2,382–2,395. https://doi.org/10.1002/asi.21641

Gleicher, N. (2019a, February 1). Taking down coordinated inauthentic behavior in Indonesia. *About Facebook.* https://about.fb.com/news/2019/01/taking-down-coordinated-inauthentic-behavior-in-indonesia/

Gleicher, N. (2019b, March 28). Removing coordinated inauthentic behavior from the Philippines. *Facebook Newsroom*. https://about.fb.com/news/2019/03/cib-from-the-philippines/

Green, K. R. (2016). People's war in cyberspace: Using China's civilian economy in the information domain. *Military Cyber Affairs*, *2*(1), p. 5.

Grimme, C., Assenmacher, D., & Adam, L. (2018). Changing Perspectives: Is It Sufficient to Detect Social Bots? In G. Meiselwitz (Ed.), *Social Computing and Social Media. User Experience and Behavior* (pp. 445–461). Springer International Publishing. https://doi.org/10.1007/978-3-319-91521-0_32

Guest, P. (2018, May 9). Malaysia elections: The inside story of Malaysia's prolific election fixer. *WIRED*. https://www.wired.co.uk/article/election-malaysia-2018-general-fake-news-day-2008-syarul-ema

Haboddin, M. (2012). Menguatnya politik identitas di ranah lokal. *Jurnal Studi Pemerintahan*, *3*(1).

Harahap, R. (2018, March 27). Two years sought for Saracen leader. *The Jakarta Post*. https://www.thejakartapost.com/news/2018/03/27/two-years-sought-for-saracen-leader.html

Harrison, S. (2019, July 7). Twitter's disinformation data dumps are helpful—To a point. *Wired*. https://www.wired.com/story/twitters-disinformation-data-dumps-helpful/

Holt, T. J., & Bossler, A. M. (2016). *Cybercrime in progress: Theory and prevention of technology-enabled offenses*. Routledge.

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, social media and political polarization in the United States, 2012–2018* (Working Paper 2018.2).

Information operations directed at Hong Kong. (2019, August 19). *Twitter*. https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html

Inter-Racial and Religious Confidence Circles (IRCCs). (2019, September 30). *Ministry of Culture, Community and Youth*. https://www.mccy.gov.sg/sector/initiatives/inter-racial-and-religious-confidence-circles

Kaiser expose: Cambridge Analytica pitched plan to UMNO on retaining 40 seats in GE14. (2020, January 3). *The Star Online*. https://www.thestar.com.my/news/nation/2020/01/03/kaiser-expose-cambridge-analytica-pitched-plan-to-umno-on-retaining-40-seats-in-ge14

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, *37*(2), pp. 256–280.

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, *111*(3), pp. 484–501.

Koh, F. (2019, March 2). Parliament: Stronger laws planned to combat foreign interference. *The Straits Times*. https://www.straitstimes.com/singapore/stronger-laws-planned-to-combat-foreign-interference

Kragh, M., & Åsberg, S. (2017). Russia's strategy for influence through public diplomacy and active measures: The Swedish case. *Journal of Strategic Studies*, *40*(6), pp. 773–816.

Lamb, K. (2017, October 1). Beware the red peril: Indonesia still fighting ghosts of communism. The Guardian. https://www.theguardian.com/world/2017/oct/01/beware-the-red-peril-indonesia-still-fighting-ghosts-of-communism

Latif, A. I. (2011). *Hearts of resilience: Singapore's community engagement programme*. Institute of Southeast Asian Studies.

Leong, P. P. Y. (2019). *Malaysian politics in the new media age: Implications on the political communication process*. Springer.

Lessenski, M. (2018). *Common sense wanted: Resilience to 'post-truth' and its predictors in the new media literacy index 2018*. Open Society Institute — Sofia.

Lewis, R., & Marwick, A. E. (2017, December 15–16). Taking the red pill: Ideological motivations for spreading online disinformation. Understanding and Addressing the Disinformation Ecosystem Workshop, University of Pennsylvania Annenberg School for Communication, Philadelphia, PA. https://pdfs.semanticscholar.org/9d91/58807cbf03fff609e74ef9e0e61c2e6088d8.pdf#page=21

Linvill, D. L., & Warren, P. L. (2020). Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication*, *37*(4), pp. 447–467.

Love, J., & Cooke, K. (2016, November 15). Google, Facebook move to restrict ads on fake news sites. *Reuters*. https://www.reuters.com/article/us-alphabet-advertising/google-says-working-on-policy-update-to-restrict-ads-on-fake-news-sites-idUSKBN1392MM

Maulia, E., & Nugroho, B. (2017, August 24). Indonesia strikes at alleged internet fake news syndicate. *Nikkei Asian Review*. https://asia.nikkei.com/Politics/Indonesia-strikes-at-alleged-internet-fake-news-syndicate

Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, *3*(1), pp. 39–43.

Mueller, R. S. (2019). *Report on the investigation into Russian interference in the 2016 presidential election* (Volume 1). U.S. Department of Justice.

Occeñola, P. (2018, February 9). From Korean idol group to Duterte supporters on Facebook? *Rappler*. https://www.rappler.com/technology/social-media/195675-korean-star-facebook-fan-group-turned-duterte-supporters

Ong, J. C., & Cabañes, J. (2018). *Architects of networked disinformation: Behind the scenes of troll accounts and fake news production in the Philippines*. Newton Tech4Dev Network.

Ong, J., Tapsell, R., & Curato, N. (2019). Tracking digital disinformation in the 2019 Philippine midterm election. *New Mandala*. www.newmandala.org/disinformation

Osman, M. N. M. (2010). The transnational network of Hizbut Tahrir Indonesia. *South East Asia Research*, *18*(4), pp. 735–755.

Packham, C. (2019, December 5). Australia to probe foreign interference through social media platforms. *CNBC*. https://www.cnbc.com/2019/12/05/reuters-america-australia-to-probe-foreign-interference-through-social-media-platforms.html

Paramaesti, C. (2018, February 28). Begini perbedaan the family MCA dengan saracen. *Tempo.Co*. https://nasional.tempo.co/read/1065322/begini-perbedaan-the-family-mca-dengan-saracen

Paul, C., & Matthews, M. (2016). The Russian "Firehose of Falsehood" propaganda model. *RAND Corporation*.

Perkins, T. (2019, July 5). Facebook ads funded by "dark money" are the right's weapon for 2020. *The Guardian*. https://www.theguardian.com/technology/2019/jul/05/facebook-ads-2020-dark-money-funding-republican-trump-weapon

Persada, S., Dongoran, H. A., & Chairunnisa, N. (2017, August 28). Bos saracen mengaku pendukung Prabowo, berikut blak-blakan. *Tempo*. https://nasional.tempo.co/read/903785/bos-saracen-mengaku-pendukung-prabowo-berikut-blak-blakan

Pomerantsev, P., & Weiss, M. (2014). *The menace of unreality: How the Kremlin weaponizes information, culture and money*. Institute of Modern Russia New York.

Ramadlan, F. S., & Masykuri, R. (2018). Marketing isu agama dalam pemilihan kepala daerah di indonesia 2015–2018. *Jurnal Penelitian Politik*, *15*(2), pp. 249–265.

Roth, Y. (2019, June 13). Information operations on Twitter: Principles, process, and disclosure. *Twitter*. https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html

Sejarah banser, barisan pemuda NU pembela bangsa. (2018, October 26). *CNN Indonesia*. https://www.cnnindonesia.com/nasional/20181025170705-20-341448/sejarah-banser-barisan-pemuda-nu-pembela-bangsa

Sinaga, D. A., Almanar, A., & Setuningsih, N. (2017, July 19). Gov't officially disbands Hizbut Tahrir Indonesia. *Jakarta Globe*. https://jakartaglobe.id/news/govt-officially-disbands-hizbut-tahrir-indonesia

Speech introducing the National Security Legislation Amendment (Espionage and Foreign Interference) Bill 2017. (2017, December 7). Malcolm Turnbull. https://www.malcolmturnbull.com.au/media/speech-introducing-the-national-security-legislation-amendment-espionage-an

Standish, R. (2017, March 1). Why is Finland able to fend off Putin's information war? *Foreign Policy*. https://foreignpolicy.com/2017/03/01/why-is-finland-able-to-fend-off-putins-information-war/

Stefano, M. D. (2019, January 14). These Facebook pages are spending thousands of pounds trying to influence your views on Brexit. *BuzzFeed*. https://www.buzzfeed.com/markdistefano/facebook-pages-brexit-donations

Stephan, W. G., Ybarra, O., & Morrison, K. R. (2009). Intergroup threat theory. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 43–59). Mahwah, NJ: Lawrence Erlbaum Associates.

Sukumaran, T. (2017, August 27). Religion, race, politics: What's causing Malaysia's great divide? *South China Morning Post*. https://www.scmp.com/week-asia/society/article/2108367/religion-race-politics-whats-causing-malaysias-great-divide

Tan, G., Wan, W. P., & Teo, J. (2014). *SURE campaign: Promoting information literacy awareness to Singaporeans*. IFLA WLIC 2014—Lyon—Libraries, Citizens, Societies: Confluence for Knowledge.

Townsend, T. (2016, November 21). The bizarre truth behind the biggest pro-Trump Facebook hoaxes. *Inc.Com*. https://www.inc.com/tess-townsend/ending-fed-trump-facebook.html

Uren, T., Thomas, E., & Wallis, J. (2019). *Tweeting through the Great Firewall: Preliminary analysis of PRC-linked information operations on the Hong Kong protests*. Australian Strategic Policy Institute.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017, May 3). Online human-bot interactions: Detection, estimation, and characterization. *Eleventh International AAAI Conference on Web and Social Media*. Eleventh International AAAI Conference on Web and Social Media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587

Weedon, J., Nuland, W., & Stamos, A. (2017). *Information operations and Facebook*. Facebook Security.

Wilcox, P., Gialopsos, B. M., & Land, K. C. (2012). Multilevel criminal opportunity. In F. T. Cullen & P. Wilcox (Eds.), *The Oxford handbook of criminological theory*. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199747238.001.0001/oxfordhb-9780199747238-e-30

Wineburg, S., & McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information. *Stanford History Education Group Working Paper No. 2017-A1*. https://dx.doi.org/10.2139/ssrn.3048994

# Section D

# Cyber Fraud and Scams

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0009

# Chapter 9

# Love Cheats: The Psychology of Love Scams

Jeffery Chin

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Jeffery_CHIN@mha.gov.sg*

## 9.1 Introduction

"This guy followed me on Instagram. He told me that his parents are Singaporean but he is born and raised in London. He said he is a pilot of Thai airways. After a few exchange of msgs in Instagram he asked for my whatsapp number. After giving my number, the next day, I realised that I couldn't find his Instagram page anymore.

After a week, he then told me that he sent me a parcel with gifts from him since i've been a REAL friend and he got selected to fly the plane with the UN Secretary after a successful interview. He insisted on sending a package containing luxury items such as Chanel handbag, jewellery, valued at GBP$15,000. He sent me a receipt of his transaction and track code.

Today, he asked me if I've received the package as his was an express courier. When I asked him to check with the courier directly, he was full of angst and told me that he wasn't in Singapore. When I went to the website to check and was told that I needed to pay S$1750 as my

> parcel was overweight. This is obviously a scam, so beware ladies when chatting with online friends!"
>
> [Anonymous, 2019]

The above excerpt provides some insights into a love scammer's modus operandi. This excerpt was published by an alert member of the public on a scam prevention website[a] in Singapore, seeking to warn members of the public about love scams.

Love scam, also known as romance scam or fraud, is an advanced fee fraud where scammers fake romantic intentions to secure the trust and affection of their victims before cheating them of their money [Whitty, 2018]. Unlike other types of scams (e.g., e-commerce, impersonation scams), love scams are highly personal, as victims are induced by the romance feature of the scam. Whereas victims would typically be motivated to maintain the romantic relationship, the objective of scammers is ultimately monetary gain. While love scams have existed prior to the advent of the internet (e.g., newspaper dating classifieds), newly invented mass communication platforms (e.g., social networking sites) are commonly exploited to perpetuate the fraudulent offence in recent years [Rege, 2009].

The phenomenon of love scams appears to be a global concern, given the numbers of cases reported internationally. Given modern-day scam's heavy reliance on mass communication media, it appears common for the offence to be prevalent in countries with relatively advanced mass communication infrastructure and high Internet penetration rates. For instance, Action Fraud, the UK's national reporting centre for fraud and cybercrime received 4555 reports of love scams in 2018, with losses of more than £50 million [City of London Police, 2019]. The Federal Trade Commission in the US documented more than 21,000 reports in 2018, with victims losing US$143 million [Fletcher, 2019]. Similar trends have been reported in Australia, with the Australian Competition and Consumer Commission estimating about 4000 cases of love scams in 2019, with losses of more than AUD28 million reported and an average loss of AUD19,000 per victim [Australian Competition and Consumer Commission, 2020]. Singapore, a cosmopolitan city-state in South-east Asia, has seen significant increases in online commercial

[a]National Crime Prevention Council (2020). Scam Alert. https://www.scamalert.sg

crimes since 2014, despite being of one of the safest cities in the world [Nair, 2017]. Singapore reported 649 cases of love scams in 2019 with $34.6 million lost that year, with the largest amount lost in a single case was $4.6 million [Singapore Police Force, 2020]. The reported figures are plausibly the tip of the iceberg, as love scams are typically underreported due to the embarrassment associated with being a victim of such an offence [King & Thomas, 2009]. There are also instances where individuals may not even realise or refuse to believe that they are victims [e.g., Alkhatib, 2018].

All the countries mentioned have significant proportions of their population connected to the internet, with more than 80% internet users in 2019 and more than 65% being active social media users [We Are Social, 2019]. Victims were often approached on popular dating and social media platforms such as Tinder, Facebook, and Instagram in Singapore [Aw, 2018; Burgess, 2019].

In view of this crime trend and its consequences, several streams of research have emerged internationally to examine the phenomenon. Drawing from the extant research literature, developments, and anecdotes from cases reported in Singapore, this chapter aims to provide an overview of the offence and insights to its mechanisms from a psychological perspective. It begins by articulating the typical structure of love scams, before elaborating on the different factors that make them possible. The impact of love scams on its victims is then discussed, noting that the emotional consequences to victims are often overlooked by the victims' support network. Recommendations on proposed measures needed to tackle this global issue are finally presented towards the end of this chapter.

## 9.2 How do Love Scams Work?

As described by Whitty [2013]'s Scammers Persuasive Technique's Model, a scammer typically initiates and builds a relationship with their victim online through dating and/or social networking sites. The scammer would profess his/her romantic intentions towards the victim early in the relationship and request for the victim to move their existing communication medium from public dating/social media sites to more immediate and exclusive forms of private communication such as emails or instant messengers [Whitty, 2013]. One plausible reason that scammers do this is to prevent administrators of the sites from detecting and closing

down their fake accounts, thus prematurely terminating their engagement with potential victims.

Intimate exchanges of information would ensue over a period of time, from which the scammer studies the victim's preferences and present themselves as an ideal partner. Through constant online grooming over a period of time, the potential victim would grow to trust the scammer, resulting in the formation of a strong attachment. During this period, scammers may at times shower their victims with gifts or provide them with fake official documents (e.g., passports, false documents about their non-existent companies). In addition, scammers may involve third parties as these actions would inadvertently enhance their sincerity and increase the authenticity of the relationship for the victims [Buchanan & Whitty, 2014].

The scammer would monitor the relationship and make monetary and/or sexual requests once a desired level of trust is attained. Such requests may either take the form of small monetary requests that increases in amount over time (i.e., 'foot in the door' technique), or in a form of a 'crisis' that happened to the scammers, a situation which they would use to convince the victims to assist them financially. Some scammers may also employ the 'door in the face' technique, a persuasive technique where scammers ask extreme favours of their victims knowing that they would refuse, before settling on moderate requests [Whitty, 2013].

Victims are continually exploited over a period of time in their bid to sustain the perceived romantic relationships. In the process, these victims may potentially experience both financial losses and sexual abuse. Some victims may unwittingly become embroiled in money laundering offences (e.g., acting as money mules) when they handle monetary transfers on behalf of their scammers. For instance, a 48-year-old woman in Singapore was reportedly so enamoured with her fake online lover that she continued to help him launder money (money from other scams victims) using her family's and her own bank accounts, despite being warned by the police [Koh, 2019].

There have also been instances where victims were sexually exploited by their scammers and extorted for large sums of money when they acceded to the request of scammers to perform sexual acts in front of the computer to prove their commitment to the romantic relationship. A series of such cases occurred in Singapore in 2011. In one of the cases, a 22-year-old man stripped naked in front of his webcam in order to prove

his sincerity in obtaining the phone number of a 17-year-old girl whom he befriended on an internet chat room. Unbeknownst to him, his exploits were recorded by the girl's boyfriend, who later extorted him of more than $97,000. The couple also targeted five other victims using similar modus operandi [Chong, 2012]. It is also evident from this instance that victims of love scams are not limited to women.

The scam process ends when victims lose patience with the relationship and/or becomes aware of the scam. It has also been documented that some victims may be re-victimised even after being scammed before. For instance, scammers may feign remorse and seek victims' forgiveness for the scam, claiming that they fell in love with their victims in the process of the scams [Whitty, 2013]. Other instances may include different scenarios where scammers pose as law enforcement officers seeking fees from victims for certain law enforcement procedures (e.g., recovery of money) related to their cases [Whitty, 2013].

## 9.3 The Internet: A Conducive Environment for Love and Crime

Relationship formation on the internet is a unique feature of modern life. Studies have established that people do form personal relationships with others on the internet [e.g., Donn & Sherman, 2002; Parks & Floyd, 1996; Parks & Roberts, 1998], including intimate relationships. In Singapore, a survey by the National Population and Talent Division in 2016 revealed that respondents (2,940 Singapore Residents between ages of 21 to 45) are increasingly open to meeting their future partners online, with the figures rising from 19% in 2012 to 43% in 2016 [Lee, 2017].

Several features of internet communication platforms (also known as computer mediated communication platforms, or CMC) render it conducive for relationship formation. McKenna, Green and Gleason [2002] posit that the *relative anonymity* and oftentimes, *perceived invisibility* (e.g., presentation of oneself through an online identity/avatar) on CMC platforms (e.g., online chatrooms, social media, email, blogs, forum, dating websites) offers users a safe environment that facilitates the disclosure of intimate information about themselves without the fear of stigma or reprisal. The process of self-disclosure is known to be an integral ingredient in relationship formation as it leads to the enhanced experience of intimacy during interactions on CMC platforms.

Relationship formation on the CMC platforms is also commonly facilitated by the absence of 'gating features' such as physical appearances, deficiencies in social skills (e.g., shyness, stuttering), and differences in social/power status [McKenna *et al*., 2002]. In the absence of such information, it is posited that CMC users may instead focus on shared interests and values to decide whether to pursue a relationship [Donn & Sherman, 2002]. The absence of gating features on the internet also offers individuals who are less attractive or less confident of their social abilities—as compared to their more attractive and socially skilled counterparts—similar opportunities to interact and develop relationships, oftentimes to a point where the disclosure of intimate information occurs.

In a similar strain, owing to the visual invisibility on many CMC platforms, users need not have to worry about their non-verbal behaviours and self-presentation skills during interaction with their partners. Instead, they are able to focus more cognitive resources towards crafting meaningful messages or replies [Suler, 2004]. This allows users the opportunities to probe in-depth into interpersonal matters, align their values and interests to that of their online partners, and in the process, enhance their desirability to their online partners [Walthers, 1996].

The *convenience* afforded by CMC platforms is also another feature that makes such platforms conducive for relationship formation [Rege, 2009]. Individuals do not need to leave the comfort of their homes or rearrange their work schedules to maintain their online relationships. In particular, asynchronous CMC platforms (i.e., emails, blogs, forums where people reply to messages at their convenience) do not require both partners' simultaneous attention and allow partners to interact at their convenience. The ease of accessibility to CMC platforms offered by various modes of communication devices in recent years (e.g., computers, mobile phones, mobile tablets) also means that individuals are able to maintain their online relationships throughout the day, making these relationships a routine, yet salient aspect of their lives [Whitty, 2013]. Communicating via such platforms also leave records for individuals to revisit at their convenience, which may serve to further reinforce the romantic messages sent by scammers [Whitty, 2013].

The same features of the CMC platforms that render it conducive for relationship formation also make it an ideal conduit for scammers to exploit for their malicious intents. The relative anonymity and invisibility offered by the internet provides scammers a secure environment to

operate in. There is a general perception among users that most individuals are not technologically savvy to the extent of tracking computer IP addresses [Suler, 2004]. Even if IP addresses are tracked, these can be easily masked with the availability of VPNs. In the event their attempts at scamming result in failure (e.g., suspicious victims), scammers can easily disengage and cut contact with minimal consequences.

Scammers tend to engage in multi-layered transactions and may use virtual currencies to bypass financial controls and avoid tracing efforts [Nair, 2017]. They take advantage of the inter-connectivity of the internet, exploit trans-jurisdictional barriers, and perpetuate their offences virtually in countries they do not reside in, so that arrest, investigation, and prosecution are challenging. For instance, a number of love scam syndicates targeting Singapore victims in recent years appear to be transnational and are often perpetrated by African gangs based in Malaysia ["KL-based fraud syndicate," 2020; Lim, 2018; Mahmud, 2018; "Singapore helps busts transnational," 2019].

On CMC platforms, individuals are also able to control the information they present, with their unsuspecting victims being often unable to check their authenticity. Scammers are known to operate under fake profiles stolen from modelling websites and at times, unknowing individuals [Rege, 2009; Whitty, 2013]. The fake profiles would tend to be aligned to the typical features that males and females seek in a partner, making them appealing to potential victims [Whitty, 2013]. For instance, it is posited by researchers that men tended to prefer partners who are physically attractive while women prefer partners with high socioeconomic statuses [Kenrick *et al*., 1990]. Research has also suggested that people tend to perceive information present on the internet to be credible [Whitty & Joinson, 2009]. Such features afforded by CMC platforms create the perception of safety that renders scammers less inhibited to engage in deviant behaviours, including committing crimes such as love scams [Suler, 2004].

Most scammers claim to reside overseas, rendering face-to-face meet-ups difficult. This provides scammers with excuses to convince their victims to interact exclusively on CMC platforms. Communication on such platforms offers scammers opportunities to study their potential victims (e.g., via the 'digital footprints' of their victims) and present themselves as attractive partners. Scammers also have time to craft and time their messages and replies in the absence of 'gating features' (e.g., non-verbal behaviours that may betray their disinterest and/or malicious intents) to gain the trust and affection of their victims.

As features of CMC platforms facilitate self-disclosure, scammers would encourage their victims to disclose intimate aspects of themselves, be attentive towards their victims' preferences, and be non-judgemental towards flaws revealed by their victims. In return, scammers would reveal intimate (but false) aspects of themselves to gain the trust of their victims. Consequently, strong attachments form between the victims and their scammers. Such attachments prove advantageous to the scammers who subsequently utilise this connection to exploit their victims for monetary gains.

## 9.4 Psychologically Persuasive Tactics

Scammers are known to take advantage of common errors in judgement and use well-known social influence tactics as they groom and exploit their victims. Muscanell, Guadagno and Murphy [2014] posit that people tend to be susceptible to such tactics on the internet as they tend to experience cognitive overload from vast volumes of information they encounter online. Moreover, people also tend to engage in several tasks online simultaneously, resulting in less effortful thinking, in addition to developing the tendency to use cognitive shortcuts to assess information and make decisions [Muscanell *et al*., 2014; Guadagno, Okdie, & Muscanell, 2013; Srivastava, 2013].

A common fake profile adopted by male scammers is one of a person of *authority* (e.g., military personnel, engineer, businessman). It is well documented that people tend to obey and trust the authorities [e.g., Milgram, 1963]. This specific error in judgement is also commonly employed in marketing advertisement tactics, where highly esteemed professionals are invited to endorse the products that these companies are selling.

*Norm activation*, where victims act according to what they believe is socially appropriate in a given situation, is another common tactic employed by scammers [Lea *et al*., 2009; Whitty, 2013]. The fake profiles presented by scammers tended to contain elements of vulnerability that would prime victims towards helping when an opportunity arises, as it activates a basic human norm to help someone who is in need [Batson, 1998]. It is expected that one would be obliged towards helping a romantic partner in need of help. Scammers would groom victims with the goal of convincing them that the romantic relationship they share is genuine, before making requests of said victims.

During the period of courtship, scammers would present themselves as being accepting of, and non-judgemental towards the flaws of their targets [Whitty & Buchanan, 2012]. They would use *flattery* and messages of endearment to woo their targets. Flattery tends to make targets feel good and like the communicator [Jones, 1964; Taylor & Brown, 1988]. In addition, scammers would engage in selective self-presentation and present themselves as being *similar* to their targets (e.g., shared interest, experiences, values). This enhances the likelihood of a strong attachment forming. Research also suggests that we tend to like people who are similar to us [Cialdini, 1984]. Owing to the *liking* created after a period of courtship, targets are more likely to comply with the requests of their scammers as individuals are more likely be persuaded by people that they like [Cialdini, 1984].

Some scammers may also exploit the *reciprocity* principle where they send small gifts to their targets during the period of courtship [Cialdini *et al*., 1975]. The gifts serve to enhance the authenticity of the relationship for their targets, thereby contributing to the likelihood of their targets complying when these scammers make their requests.

The requests of scammers can be in the form of small requests that increases in size over time or an extreme request (in the guise of a crisis) which may be subsequently moderated. The former is a well-documented persuasion technique known as 'the foot in the door' technique [Freedman & Fraser, 1966]. This technique is known to work well in influencing individuals to comply in situations where the request is pro-social in nature (e.g., helping a loved one in need) [Beaman, Cole, Klentz, & Stenblay, 1983].

The latter is known as 'the door in the face' technique and common explanations for how it works are linked to reciprocity concerns and/or norms activation [Cialdini *et al*., 1975; Tusing & Dillard, 2000; Turner, Tamborini, Limon, & Zuckerman-Hyman, 2007]. The rejection of the first request compels victims to accede with the second, less extreme request as it would be perceived that their scammers have compromised on their initial requests. Consequently, targets would be compelled to reciprocate to the compromise. The activated social norm of helping less fortunate others may also compel targets to accede to the more moderate request.

The 'crisis' created by scammers may also activate the principle of *scarcity* where targets would be persuaded to comply with the requests of scammers. It is plausible that the crisis presented would be perceived as a rare event and targets would have little time to make informed decisions

before they comply to the requests [Lea *et al*., 2009; Muscanell, Guadagno, & Murphy, 2014].

Victims are known to be continually exploited by their scammers until such a time where they lose patience with the relationship or uncover the scam. The *sunk cost* effect is likely to be a factor at work that explains for repeated victimisation. Individuals are known to have a strong desire to be consistent and can be persuaded to act consistently [Cialdini, 1984]. Victims may continually be victimised as they may perceive that since they have invested so much into the relationship, acceding with the next request may fulfil their wish of a successful relationship [Lea *et al*., 2009].

## 9.5 Risk Factors the Render Victims Vulnerable

Whitty [2018] highlighted that victims of love scams tended to be women who are middle-aged and well educated. Her findings mirror that of Singapore, where victims are reported to be mostly middle-aged women of varying marital statuses and professions [Burgess; 2019; Lee, 2016; Mahmud, 2018].

In her study, Whitty [2018] reported that in terms of personality characteristics, victims tended to be more impulsive, trusting of others, and are pre-disposed towards addiction. In an earlier study, Whitty [2013] reported that victims are (1) typically individuals who are motivated by the hope of finding a romantic partner online; and (2) a number of them may have been searching for a period of time or have recently underwent a break-up. It has been posited that victims of love scams may have been motivated to date online by loneliness, and possess certain personality characteristics (e.g., higher in Agreeableness, Extraversion, lower in Neuroticism) that render them vulnerable [Buchanan & Whitty, 2014]. In addition, Buchanan and Whitty [2014] found that victims of love scams tended to idealise their romantic partners and experience romantic relationships intensely. This may lead them to be less sensitive to red flags that might suggest that their online partners are scammers.

Fisher, Lea, and Evans [2013] suggested that it is possible that some scam victims may have been overconfident in their abilities to detect scams, which may explain why well-educated individuals fall prey to scams. Debatin, Lovejoy, Horn, and Hughes [2009] similarly posit that people generally tended to think that others, but not themselves, are more

likely to fall prey to online scams. Low [2019] reported findings from a study on love and impersonation scams conducted by the Singapore Police Force. The study found that victims tended to have an "optimism bias" that made them believe that they are safe from scams, as they think that Singapore is a relatively safe country with low crime rates, and that they are confident in their abilities to protect themselves.

## 9.6 Scam Prevention and Interventions

### 9.6.1 *Public education on love scams*

Given the prevalence and consequences of online love scams, it is pertinent that the public is made aware of this offence, so that potential victims may be prevented from falling prey. In particular, scams prevention efforts should be targeted at users of social media platforms and dating websites, as it is commonly on such platforms where victims are first targeted. Users of such platforms can be introduced to trusted online resources (e.g., scam prevention websites such as *scamalert.sg, romancescams.org*) to sensitise them to the common modus operandi of online love scammers and be made aware of the contact channels of the various support services currently available.

Some of these anti-scams sites also offer databases containing common fake profiles used by scammers (e.g., photos, emails, and telephone numbers) where online daters can verify details of their online partners against. Online daters can also be introduced to simple resources that they may use to verify against the details of their online partners. For instance, Google, an online search engine, allows users to conduct searches using existing images available on the web. Users can be encouraged to reverse search images of their online partners using such software to assess the authenticity of the photos used by their online partners.

It is noteworthy that Cross and Kelly [2016] cautions against the provision of too much information on scams (e.g., detailed typology of different types of scams, different MOs of scammers) in public education efforts, as it risks information overload and distracts the public from the key messages. Instead, the authors claim that the messaging around such efforts should be clear, simple, and be focused on how the public should protect their money and personal details regardless of the types of scams or MOs they may encounter.

### 9.6.2 *Automatic detection of fake scam profiles*

Given the volume of information that users engage with on the internet, the sophisticated social influence tactics used, and the common decision-making errors exploited by scammers, it may be difficult for users of dating and social media platforms to determine if the person they are interacting with is a scammer. It is arguable that technical solutions are required to complement public education efforts.

Major social media companies have explored the use of machine learning based content detection techniques to automatically detect and remove extremist content from their platforms ["Facebook, YouTube, Twitter and Microsoft join," 2017]. Using similar techniques, Suarez-Tangil and colleagues [2019] successfully tested a prototype system capable of automatically detecting fake scam profiles at promising accuracy rates of 97% [Suarez-Tangil *et al*., 2019]. With further development, such systems have the potential to be introduced to dating platforms to supplement existing scam prevention efforts, further enhancing protection for users of such platforms.

### 9.6.3 *Involve industrial partners in crime prevention*

Rege [2009] highlighted that the online dating industry is rarely regulated, and many dating websites offer few protection measures for their users against scammers. Often, many dating websites take minimal measures to protect their users, by only asking users to surface 'problematic partners' through 'report abuse' features. While such websites have increasingly posted advisories to warn their users to exercise caution, such measures may be insufficient as users may get habituated after repeated use and ignore such advisories over time.

Authorities may consider regulating the local online dating industry on a basic level, by mandating that local companies hosting dating websites implement certain protection features for their users. For instance, users of such websites may be asked for certain identification details (e.g., credit card details) when they register. In addition, the companies may be required to proactively monitor the online behaviours of their users and to block users with unusual behaviours. Such companies may also be required to provide access to support resources (e.g., scam education) and establish clear incident reporting channels for potential victims of scams.

Financial institutions involved in money transfers (e.g., banks, money remittance companies) can also be involved in crime prevention efforts. For instance, the Singapore Police Force set up an Anti-Scam Centre in June 2019 to centralise and streamline efforts to tackle online scams. One key initiative of the Centre involves working closely with major local banks to enhance the monitoring and freezing of suspicious banks accounts and transactions. The initiative is reported to have enabled the recovery of significant amounts of money that was lost to scams in mid-2019 [Ng, 2019].

### 9.6.4 *Educate support networks about the significant emotional distress experienced by scam victims*

Prior research has documented the emotional and psychological impacts of victimisation on scam victims [Buchanan & Whitty, 2014; Cross, Richards, & Smith, 2016; Modic & Anderson, 2016; Whitty & Buchanan, 2016]. Particular to love scams, victims often experience two key losses: financial and relationship [Whitty & Buchanan, 2012; Buchanan & Whitty, 2014; Whitty & Buchanan, 2016]. In their study, Whitty and Buchanan [2016] found many victims to be more distressed by the loss of a significant relationship rather than the financial loss. In fact, the emotional experience for some victims was described to be akin to grieving for the loss of a loved one. Victims described experiencing a host of negative emotions, including anger, shame, guilt, fear, and depression, and some individuals reported being suicidal, and becoming distrustful of people and unable to form new romantic relationships.

Members of their social network may berate the victims and be frustrated as some of them may have attempted to warn the victims to no avail, leaving victims with potentially little support [Button, Lewis, & Tapley, 2014; Cross, Richards, & Smith, 2016; Whitty & Buchanan, 2016]. Similarly, insensitive treatment by law enforcement officers may result in further trauma for such victims, which may in turn, deter other victims from coming forward [Whitty & Buchanan, 2016].

Hence, it is crucial to educate the public on the plight of scam victims and facilitate victims' access to psychological support services as early as possible. Given the significant level of distress experienced, it may also be necessary for love scam victims to be treated as vulnerable victims with special measures accorded to them by law enforcement and the courts [Whitty & Buchanan, 2016].

## 9.7 Conclusion

Despite its benefits, threats abound on the internet. Online scams are one of these threats. As more people across the world are connected to the internet, there is an urgent need for solutions to protect them from these threats. Love scams, in particular, appear to be a global concern in recent years. While there has been much more effort invested in public outreach and education, research should focus on identifying effective public education strategies that protect the public against the scourge of scams. Potential research areas include how 'nudging' principles can be incorporated into scam prevention messages to enhance its effectiveness and whether experiential learning approaches may be effective in scam prevention initiatives to counteract the effects of optimism bias towards scams in some individuals. In addition, there needs to be a broader effort at enhancing levels of cyber hygiene, so that the public can protect themselves against a wider range of cyber threats, including scams. Lastly, given the rapid advancement in information technology, the volume of information online that impacts peoples' information processing abilities and the ever-evolving, sophisticated tactics used by scammers, technological solutions need to be developed to supplement public education efforts, so as to further 'target harden' the internet environment for the safety of users.

## 9.8 Acknowledgement

The views expressed in this chapter are the author's alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 9.9 References

Australian Competition and Consumer Commission. (2020, February, 9). *Romance scammers move to new apps, costing Aussies more than $28.6 million.* https://www.accc.govau/media-release/romance-scammers-move-to-new-apps-costing-aussies-more-than-286-million

Alkhatib, S. (2018, July, 12). Despite warning, she continued to receive money from 'online lover'. *The New Paper.* https://www.tnp.sg/news/singapore/despite-warning-she-continued-receive-money-online-lover

Anonymous. (2019, October, 3). Re: I WAS ASKED TO PAY MORE THAN $1700 FOR MY GIFTS. [Blog post]. https://www.scamalert.sg/stories-details/Story-03Oct2019140710PM

Aw, C. W. (2018, March, 26). Fewer online love scams after police set up task force to sniff out fake Romeos. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/fewer-online-love-scams-after-police-set-up-taskforce-to-sniff-out-fake

Batson, C. D. (1988). Altruism and prosocial behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 282–316). Boston: McGraw-Hill.

Beaman, A. L., Cole C. M., Klentz, B., & Stenblay, N. M. (1983). Fifteen years of the foot-in-the-door research: A meta-analysis. *Personality and Social Psychology Bulletin, 9*, pp. 181–196. https://doi.org/10.1177%2F0146167283092002

Buchanan, T. & Whitty, M. T. (2014). The online dating romance scam: Causes and consequences of victimhood. *Psychology, Crime and Law, 20*(3), pp. 261–283. https://doi.org/10.1080/1068316X.2013.772180

Burgess, D. (2019, September, 5). Internet love scam victim warns others. *The New Paper.* https://www.tnp.sg/news/singapore/internet-love-scam-victim-warns-others

Button, M., Lewis, C., & Tapley, J. (2014). Not a victimless crime: The impact of fraud on individual victims and their families. *Security Journal*, *27*(1), pp. 36–54. https://doi.org/10.1057/sj.2012.11

Cialdini, R. B. (1984). *Influence: The psychology of persuasion.* New York: William Morrow.

Cialdini, R. B., Vincent, J. E., Lewis, S. K., Catalan, J., Wheeler, D., & Darby, B. L. (1975). Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology*, *31*, pp. 206–215. https://psycnet.apa.org/doi/10.1037/h0076284

City of London Police. (2019, February, 12). *Don't invest your heart in a fauxmance: victims lose over* £50 *million to romance fraud.* https://news.cityoflondon.police.uk/r/1191/don_t_invest_your_heart_in_a_fauxmance_victims_l

Chong, E. (2012, March 1). Naked video-chat scam couple charged with extortion. *The Straits Times*, p. A10.

Cross, C. & Kelly, M. (2016). The problem of 'white noise': Examining current prevention approaches to online fraud, *Journal of Financial Crime*, *23*(4), pp. 806–818 http://dx.doi.org/10.1108/JFC-12-2015-0069

Cross, C., Richards, K., & Smith, R. G. (2016). The reporting experiences and support needs of victims of online fraud. *Trends & Issues in Crime and Criminal Justice*, *518*, pp. 1–14.

Cukier, W. & Levin, A. (2009). Internet fraud and cyber crime. In F. Schmalleger & M. Pittaro. (Eds.), *Crimes of the Internet* (pp. 251–279). New Jersey: Pearson Prentice Hall.

Debatin, B., Lovejoy, J. P., Horn, A., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal*

*of Computer-Mediated Communication*, *15*, pp. 83–108. https://doi.org/10.1111/j.1083-6101.2009.01494.x

Donn, J. E. & Sherman R. C. (2002). Attitudes and practices regarding the formation of romantic relationships on the Internet. *Cyber Psychology and Behavior*, *5*(2), pp. 107–123. https://doi.org/10.1089/109493102753770499

Facebook, YouTube, Twitter and Microsoft join to fight against terrorist content (2017, June, 26). *CNBC.* https://www.cnbc.com/2017/06/26/social-media-companies-join-to-fight-against-terrorist-content.html

Fletcher, E. (2019, February, 12). Romance scams rank number one on total reported losses. *Federal Trade Commission.* https://www.ftc.gov/news-events/blogs/data-spotlight/2019/02/romance-scams-rank-number-one-total-reported-losses#end1

Freedman, J. L. & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*, pp. 195–202. https://doi.org/10.1037/h0023552

Guadagno, R. E., Muscanell, N. L., Roberts, N. R., & Rice, L. M. (2013). Social influence online: The impact of social validation and likeability on compliance. *Psychology of Popular Media Culture*, *2*, pp. 51–60. https://doi.org/10.1037/a0030592

Jones, E. E. (1964). *Ingratiation.* New York: Appleton-Century Crofts.

Kenrick, D. T., Sadalla, E. K., Groth, G., & Torst, M. R. (1990). Evolution, traits and the stages of human courtship: Qualifying the parental investment model, *Journal of Personality*, *58*, pp. 97–116. https://doi.org/10.1111/j.1467-6494.1990.tb00909.x

King, A. & Thomas, J. (2009). You can't cheat an honest man: Making ($$$ and) sense of the Nigerian E-mail scams. In F. Schmalleger & M. Pittaro. (Eds.), *Crimes of the Internet* (pp. 206–224). New Jersey: Pearson Prentice Hall.

KL-based fraud syndicate that cheated Singapore Internet love scam victims of $700k busted. (2020, February, 18). *The StraitsTimes.* https://www.straitstimes.com/singapore/courts-crime/kl-based-fraud-syndicate-that-cheated-singapore-internet-love-scam-victims-of

Koh, W. T. (2019, June, 19). Woman who laundered scam proceeds for online lover jailed 40 months. *Yahoo News Singapore.* https://sg.news.yahoo.com/woman-who-laundered-scam-proceeds-for-online-lover-jailed-40-months-114551894.html

Lea, S., Fischer, P., & Evans, K. (2009). The psychology of scams: Provoking and committing errors of judgement. *Report for the Office of Fair Trading*. https://webarchive.nationalarchives.gov.uk/20140402205717/http://oft.gov.uk/shared_oft/reports/consumer_protection/oft1070.pdf

Lee, D. (2016, June, 11). Beware the Cyber Lover and His Ploys. *Home Team News.* https://www.mha.gov.sg/hometeamnews/our-community/ViewArticle/cyber-lover-and-his-ploys

Lee, S. X. (2017, July, 8). More Singaporeans finding love online: Marriage and Parenthood survey 2016. *The Straits Times*. https://www.straitstimes.com/singapore/more-singaporeans-finding-love-online-government-survey

Lim, S. (2018, October, 23). Love scam syndicate worth nearly S$20 million crippled by authorities in Singapore, Malaysia and Hong Kong. *Business Insider Singapore.* Retrieved from https://www.businessinsider.sg/love-scam-syndicate-worth-nearly-s20-million-crippled-by-authorities-in-singapore-malaysia-and-hong-kong

Low, Y. J. (2019, July, 10). Think you'll never be a scam victim? Chances are you're more likely to fall for one. *Channel News Asia.* https://www.todayonline.com/singapore/think-youll-never-be-scam-victim-chances-are-youre-more-likely-fall-one

Mahmud, A. H. (2018, March, 26). Internet love scams: 'I got carried away with his words' says victim as police strengthen task force. *Channel News Asia.* https://www.channelnewsasia.com/news/singapore/internet-love-scams-police-transnational-commercial-crime-mule-10077380

McKenna, K. Y. A., Green, A. S., & Gleason, M. E. J. (2002). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues, 58*(1), pp. 9–31. https://doi.org/10.1111/1540-4560.00246

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371–378. https://doi.org/10.1037/h0040525

Modic, D. & Anderson, R. (2015). It's All Over but the Crying: The Emotional and Financial Impact of Internet Fraud. *Ieee Security & Privacy, 13*(5), pp. 99–103. doi: https://doi.org/10.1109/MSP.2015.107

Muscanell, N. L., Guadagno, R. E., & Murphy, S. (2014). Weapons of influence misused: A social influence analysis of why people fall prey to internet scams. *Social and Personality Psychology Compass*, *8*(7), pp. 388–396. https://doi.org/10.1111/spc3.12115

Nair, H. K. (2017 September). *Challenges and Opportunities of Prosecuting in the Digital Age.* Remarks made at 22nd International Association of Prosecutors Annual Conference, Singapore. https://www.agc.gov.sg/docs/default-source/newsroom-doucments/Speeches/paper-presented-by-deputy-attorney-general-hri-kumar-nair-s-c-at-the-22nd-iap-annual-conference.pdf

Ng, C. (2019, August, 30). Police set up anti scam centre, suspicious bank accounts can now be frozen in a few days. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/police-sets-up-anti-scam-centre-working-with-banks-to-disrupt-scammers

Parks, M. R. & Floyd, K. (1996). Making friends in cyberspace. *Journal of Communication*, *46*, 80–97. https://doi.org/10.1111/j.1083-6101.1996.tb00176.x

Parks, M. & Roberts, L. (1998). Making moosic: The development of personal relationships on line and a comparison to their off-line counterparts. *Journal*

*of Social and Personal Relationships*, *15*(4), pp. 517–537. https://doi.org/10.1177%2F0265407598154005

Rege, A. (2009). What's love got to do with it? Exploring online dating scams and identity fraud. *International Journal of Cyber Criminology*, *3*(2), pp. 494–512.

Ross, S. & Smith, R. G. (2011). Risk factors for advance fee fraud victimisation. *Trends & Issues in Crime and Criminal Justice*, No. 420.

Singapore Police Force (2020, 5 February). *Annual Crime Brief 2019.* https://www.police.gov.sg/media-room/statistics

Singapore helps bust transnational love scam syndicate that conned 139 people. (2019, November, 28). *Channel News Asia*. https://www.channelnewsasia.com/news/singapore/internet-love-scam-syndicate-singapore-malaysia-macau-hong-kong-12134454

Srivastava, J. (2013). Media multitasking performance: Role of message relevance and formatting cues in online environments. *Computers in Human Behavior, 29*, pp. 888–895. https://doi.org/10.1016/j.chb.2012.12.023

Suarez-Tangil, G., Edwards, M., Peersman, C., Stringhini, G., Rashid, A., & Whitty, M. T. (2019). Automatically Dismantling Online Dating Fraud. *IEEE Transactions on Information Forensics and Security*, *15*, pp. 1,128–1,137. https://doi.org/10.1109/TIFS.2019.2930479

Suler, J. (2004). The online disinhibition effect. *Cyber Psychology and Behavior, 7*(3), pp. 321–326. https://doi.org/10.1089/1094931041291295

Taylor, S. E. & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), pp. 193–210. https://doi.org/10.1037/0033-2909.103.2.193

Turner, M. M., Tamborini, R., Limon, M. S., & Zuckerman-Hyman, C. (2007). The moderators and mediators of door-in-the-face requests: Is it a negotiation or helping experience? *Communications Monograph*, *74*(3), pp. 333–356. https://doi.org/10.1080/03637750701543469

Tusing, K. J. & Dillard, J. P. (2000). The psychological reality of the door-in-the-face: It's helping, not bargaining. Journal of Language and Social Psychology, *19*, pp. 2–25. https://doi.org/10.1177%2F0261927X00019001001

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research*, *23*, pp. 3–43. https://doi.org/10.1177%2F009365096023001001

We Are Social. (2019). *Global Digital Report 2019.* https://wearesocial.com/global-digital-report-2019

Whitty, M. T. (2013). The scammers persuasive techniques model: Development of a stage model to explain the online dating romance scam, *British Journal of Criminology*, *53*, pp. 665–684. https://doi.org/10.1093/bjc/azt009

Whitty, M. T. (2018). Do You Love Me? Psychological Characteristics of Romance Scam Victims. *Cyberpsychology, Behaviour, and Social Networking*, *21*(2), pp. 105–109. https://doi.org/10.1089/cyber.2016.0729

Whitty, M. T. & Buchanan, T. (2012). The online romance scam: A serious cybercrime. *Cyberpsychology, Behavior, and Social Networking*, *15*(3), pp. 181–183. https://doi.org/10.1089/cyber.2011.0352

Whitty, M. T. & Buchanan, T. (2016). The online dating romance scam: The psychological impact on victims—both financial and non-financial. *Criminology & Criminal Justice*, *16*(2), pp. 176–194. https://doi.org/10.1177/1748895815603773

Whitty, M. T. & Joinson, A. N. (2009). *Truth, lies and trust on the Internet.* London: Routledge, Psychology Press.

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0010

# Chapter 10

# Cybercrime and Scams Amidst COVID-19: A Review of the Human Vulnerabilities Exploited During a Global Pandemic

Afreen Chawla*, John Yu†, and Shannon Ng‡

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Afreen_CHAWLA@mha.gov.sg

†John_yu_from.tp@mha.gov.sg

‡Shannon_NG@mha.gov.sg

## 10.1 Introduction

The global outbreak of the Coronavirus Disease 2019 (COVID-19) has shown little sign of abatement. As of 1 June 2020, there are more than six million confirmed cases throughout the world and these figures continue to rise [Government of Switzerland, 2020a]. The effects of this epidemic have rippled across a multitude of industries and nations as the world braces for disruption and change amidst the gloomy economic forecast and public health anxieties.

Within these uncertain times, a band of unscrupulous individuals have sought to turn these challenges into opportunities—scammers and cybercriminals. Over the past few months, an increasing number of

reports and advisories have surfaced around the world describing new and recycled variants of scams or cybercrimes that have surfaced with the outbreak [Chan *et al*., 2020]. These often appear in the form of impersonation scams, e-commerce scams, investment scams, fraudulent donations, phishing scams, malware attacks, and fake news, often with the spurious use of the term "coronavirus" or "COVID-19."

This surge in cyber criminality is not surprising, for there is a trend of such activities thriving at the onset and during major events that capture the attention and concern of many. For example, similar scams and malware dissemination campaigns have surfaced during high profile events like the Boston Marathon bombings in 2013, the royal wedding of Prince Harry and Megan Markle in 2018, and natural disasters like the Australian bushfire crisis between 2019 to 2020 [Homeland Security Today, 2020; Sarraf, 2020]. As such, the more salient the events are to a wider global audience, the more individuals are likely to attend to, or interact with these mediums that increase their risk of victimisation.

Crucially, "infodemics," or excessive amounts of new information tend to accompany these crises, usually appearing on social media and online websites. This has become an added challenge for the public to navigate, as they sieve through misinformation and deceptive communications in an attempt to differentiate facts and genuine correspondences [Gharib, 2020]. Therefore, it is important to take note of the opportunistic fraudsters and threat actors who take advantage of such situations to conduct contextualised, random, or targeted attacks among a wider and more susceptible victim population.

As nations gear up to combat and contain the COVID-19 situation, the need to protect online communities from scams and cyber-attacks should not be neglected. It is vital to understand the methods cyber criminals employ in their exploits and why people would succumb to their schemes. Policies and practices can then be better informed to protect communities and industries from cybercrimes. Therefore, this chapter seeks to consolidate the known tactics employed by cybercriminals during the COVID-19 crisis (at the time of writing). Additionally, this chapter highlights some of the fundamental human attributes in individuals that may increase their risk of victimisation. In particular, the major vulnerabilities we consider are our fear, greed, and altruistic responses during the COVID-19 climate, along with the categories of cybercrimes and scam types that correspond with them most closely. Lastly, this chapter also proposes some recommendations to combat falling prey to

such COVID-19 related cybercrimes. Suggested implementations are thus meant to address both individual and community needs for better cyber hygiene practices and multi-party collaboration.

## 10.2 Our Fear Causes Us to Be Susceptible to Cybercrime

One of the most common responses to the COVID-19 pandemic is that of widespread fear. Fear is an emotional response elicited by the perception of harm or danger (real or imagined) [Izard, 1991], which can lead to powerful physiological and behavioural changes. Fear-enabled decision-making tends to be more pessimistic, nudging individuals towards more risk-averse options [Lerner & Kelter, 2001]. While it allows for coping with imminent threats, fear can also render people vulnerable to misjudgements and decision-making errors, as their attention is focused on mitigating the biggest threat to their well-being [Rotfeld, 1988]. This is where scams come in, because our natural reaction driven by fear increases our susceptibility to cybercrime.

Scammers and cybercriminals exploit people's fears. As people are fixated on alleviating any harm to themselves or their loved ones, they tend to discount more common but relatively minor-seeming threats [Slovic, 2010], like scams. Consequently, people may easily fall victim to such scams, having failed to recognise red flags and markers of fraudulent claims. Scammers, aware of this very human tendency, take advantage of it in times of disease outbreaks (e.g., COVID-19, Ebola, SARS) and global disasters (e.g., 2019–2020 Australian bush fires) as discussed in preceding sections [Homeland Security Today, 2020; Gharib, 2020; Lu, 2015; Sarraf, 2020].

To understand what makes this fear response particularly impactful during COVID-19, we propose three reasons—(1) uncertainty; (2) loss of control; and (3) law-abiding tendency—qualities that cybercriminals and scammers prey on.

### 10.2.1 *Fear: Uncertainty*

With the coronavirus disease 2019 posing a direct, ongoing, and ever-evolving threat to one's well-being, people are seeking a permanent solution that does not yet exist. The next best alternative is to arm oneself with an arsenal of information for greater awareness and protection.

Especially during times like these, knowledge is power. To support this growing need to be informed, there is an abundance of official and unofficial information disseminated across many communication channels, covering everything from the details of infected individuals, to the extent of the disease's spread globally, to preventive measures, and more. However, alongside such useful information, there are also conflicting information and fake news floating around. Such conflicting information inevitably perpetuates an environment filled with uncertainty and confusion. This is particularly problematic when coupled with the lack of consensus on the best coping strategies for COVID-19, as the usual way of life is subsumed by foreign practices like quarantine, social distancing, and isolation which varies from country to country, culture to culture etc.

Scammers and cybercriminals, in turn, capitalise on uncertainty spawned by this new norm by creating and spreading news of fake treatments and promising the provision of important educational tools. While some cybercriminals capitalise on this opportunity to further political agendas meant to generate doubt and distrust in the local government and authorities [Gharib, 2020], others merely use the opportunity to create more targeted scams, such as the following:

#### 10.2.1.1 *Fake news*

Governments that have maintained transparency in communication, quick response times and presented a unified front have had the advantage of eliciting public support and continued belief in their strategies. However, countries facing higher scam threats may be those that have few official communication channels or a general lack of guidance from political authorities to cope with the mass disease outbreak. As a result, false statements and conspiracy theories using unsubstantiated statements such as "the virus is a man-made bio-weapon" are making the rounds [Gharib, 2020]. Consequently, there are people who do believe such statements and share them publicly, so with greater spread of misinformation, uncertainty and distrust is elevated. Such claims may be relying on the persuasion tactic of scarcity [Cialdini, 1984]; as people place increased value on commodities perceived as limited (whether in terms of quantity or quality), exclusive access to information can become a sought-after resource. As such, the dearth of go-to verified sources for reliable information, especially during the

early stages of the disease's outbreak, may have made it difficult to ascertain the reliability of new information.

#### 10.2.1.2 *Phony websites*

The intentional targeting of individuals looking for COVID-19 treatments and information is substantiated by the abundance of new website domains registered with names related to "coronavirus." For example, the unsecured website "vaccinecovid-19.com" was created on 11 February 2020, in Russia, offering the "best and fastest test for Coronavirus detection at the fantastic price of 19,000 Russian rubles" (about S$362) alongside a live heat map of infected areas [Venkat, 2020]. This website, like many others that have popped up in the wake of COVID-19, is designed to entice individuals seeking ways to mitigate their risk of infection. Unfortunately, vulnerable individuals acquiesce to the site's enticing offers and lose personal details or money to scammers.

#### 10.2.1.3 *Phishing scams*

In addition to independent fraudulent websites, a large population of scammers have been adapting their tried-and-tested methods to befit the times. Phishing, a method used by cybercriminals to deceitfully obtain personal and financial information (e.g., login details, bank account and credit card numbers), has become commonplace. However, phishing emails are now incorporating URLs and web links with "coronavirus" related terms such as "help," "victim," "relief" and "recover" [Somesh, 2020]. As sources of authority are mimicked, scammers are able to gain the trust of individuals who are, in turn, more easily convinced to comply with advanced courses of action [Cialdini, 1984], thus resulting in greater scam victimisation.

Often, malicious scammers who pose as the World Health Organization (WHO) or the Centres for Disease Control and Prevention (CDC), spoof their official domains (i.e., sender address is listed as "who.org" instead of "who.int"; "cdcgov.org" instead of "cdc.gov") and send out mass spam emails [Government of Switzerland, 2020b; ICE71, 2020]. These emails contain links for 'more information' (Click on this link to get an updated list of new cases around your city!) which upon clicking, infect the victim's devices with software that enables third-party access to personal information. Alternatively, clicking on some

URLs land individuals on fake Microsoft Outlook login pages that prompt users to enter their login credentials which are recorded by scammers to access victims' email accounts and steal other valuable information. Such malicious emails first appeared in February this year, quickly becoming a part of the first big phishing campaigns of 2020 [Weisbaum, 2020].

#### 10.2.1.4 *Malware dissemination*

The dissemination of malware is another lucrative business for scammers. Like phishing scams, malware scams also involve the use of emails that contain malicious links to automatic downloads. These contain malware software which can, if opened or downloaded, damage the computer system, and/or read, edit, or steal information. User actions can also trigger the installation of hidden software that enables unauthorised access to user devices. Besides malware URLs, scammers have also sent out emails with malicious .pdf, .mp4 and .docx attachments that contain Trojans or worms (e.g., 'Worm.Python.Agent.c', 'UDS:DangerousObject.Multi.Generic', & 'Trojan.WinLNK.Agent.gg') with malware disguised as educational documents [Buxton, 2020] to invite interest.

One such malware scam is the Emotet attack, where a Trojan (Emotet) disguised as banking software, is able to steal financial information, spy on the victims, and modify or copy personal data [Tham, 2020]. This scam has been the most active in Japan since January 2020 [Venkat, 2020]. Another instance has been a phishing campaign with the sender posing as the WHO Director-General and sharing HawkEye malware payloads designed to steal information and further fetch other malware (allowing for third-party cybercrime actors) onto unsuspecting victims' devices [Gatlan, 2020].

### 10.2.2 *Fear: Loss of control*

Individuals tend to believe either that they have control over the outcomes of events in their lives (internal locus of control), or that outside forces do (external locus of control). Those with an external locus of control are more responsive to stressful experiences and prone to greater psychological problems [Rotter, 1954]. Moreover, during threatening situations (e.g., the COVID-19 outbreak), most feel low in individual

control and high in external agency, which for many, amounts to high dread risk (perceived lack of individual control) [McDaniels *et al*., 1997; Slovic, 1987]. This loss of control fuels anxiety, panic, and irrational behaviour, manifested in recent times as panic-buying of food, toilet paper, disinfectants, etc. As manufacturers struggle to meet consumers' rising demands manifested by rapidly depleting stocks in physical stores, many seek out online avenues for buying healthcare essentials aimed at mitigating their risk of infection, such as: surgical masks, hand sanitisers and alcohol wipes. In an attempt to regain control over the spiralling events, individuals may thus lose out to scammers who exploit this vulnerability.

#### 10.2.2.1 *E-Commerce scam*

The sale of fake products (particularly surgical masks) and other e-commerce scams have been a trend across many countries experiencing the disease outbreak. Such scam campaigns are highly attractive, as masks are ostensibly the "first line of defence" against COVID-19 [Vergelis, 2020] and not freely available. Thus, utilising the persuasion tactic of scarcity [Cialdini, 1984] allows the scam to be highly successful. Advertised as authentic on legitimate-looking websites or e-commerce listings and sold at steep prices, such scams are orchestrated by price gougers and swindlers who either deliver incorrect or substandard products, if they deliver at all. Requests for refunds and returns are usually met with silence, and victims often end up worse off than when they started.

An instance of this is the fake coronavirus mask scheme in Hong Kong, a country targeted due to an early shortage of surgical masks. Orchestrated via Facebook advertisements for Japan-sourced masks priced significantly below street value [Emen, 2020], this face mask scam has resulted in a reported loss of HK$150,000 (about S$26,840); victims transferred funds as payment but never received their orders. Similarly, at least 31 million illegal, fake, and substandard face masks (worth S$35 million) have been produced for sale on the China black market and seized by the Sichuan provincial authorities ["China has Seized," 2020], incurring large losses for a population already under siege by COVID-19.

Locally, Singaporeans were faced with mask scams as well, particularly through the popular e-commerce platform 'Carousell'.

While some were intentionally scammed out of their money, e.g., by teenagers who racked up almost S$10,000 from victims [Ang, 2020], others faced losses when their own reselling attempts were thwarted by international scam suppliers (in one case, with both the new buyers and the reseller losing over S$122,000) [Chia, 2020]. Other cases mirror those witnessed internationally, as with new (sham) organisations like MedicalLex emerging to supposedly shoulder the demand of protective face masks [Zhang, 2020]. Using sunk cost tactics, e.g., creating a fake queue for people to invest their time and efforts in, companies like these have been able to sell highly overpriced masks, subsequently providing excuses to both account for "delays" and ultimately, for their failure to deliver. Commonly, individuals fall for this trap because they irrationally justify such activities as dissonant from their expectations, due to previously invested resources (effort, time, then money) [Arkes & Blumer, 1985].

Another hoax product, potentially more insidious because of its source of mass publicity, may be the Silver Solution, a "cure" for the virus which was marketed and broadcast on television by US televangelist Jim Bakker [Schwartz, 2020]. This kind of fraudulent peddling also preys on desperation stemming from the need to regain control (through a solution) over the situation.

### 10.2.3 *Fear: Law-abiding tendency*

Besides internal motivations, external factors may also play a part in fear responses. Most individuals are fearful of the consequences of non-compliance with officials or authorities, especially during a global crisis like COVID-19. Since the use of an authoritative persona lends credibility and suggests underlying expertise to claims and requests, individuals are more easily convinced to conform [Cialdini, 1984]. Unfortunately, many cybercriminals latch onto this vulnerability and fraudulently sport impressive designations (impersonate government officials, law enforcement, medical professionals, and scientists) and act as though representing well-known organisations in their mission to assuage the impact of the COVID-19 outbreak. Susceptible individuals, in a bid to abide by the requests, instructions, or demands of perceived officials may thus provide personal details, send out money, or provide credentials to scammers (and to also avoid any possible negative repercussions).

#### 10.2.3.1 *Smishing & Phishing scams* (*revisited*)

As aforementioned, several phishing emails employ the use of phony CDC and WHO credentials, spoofing their email domains (e.g., cdcgov.org instead of cdc.gov) and using official agency logos (to add credibility) while keeping their messaging short and easy to follow. This tends to work also because of the use of attention-grabbing content (e.g., subjects titled "Coronavirus outbreak in your city (Emergency)"). The content itself appeals to present public concerns [Weisbaum, 2020], while the use of authority as a persuasion tactic warrants compliance [Cialdini, 1984]. Scammers have so far posed as doctors (Specialist Wuhan-virus-advisory), officials (CDC-INFO National Contact Centre) and in more extreme cases, even the president [Letter from President] [Holmes, 2020].

While some such emails are sent out directly, others are sent via "smishing" or SMS phishing, where recipients are prompted to click on fake alert warnings (e.g., outbreak in "Back Bay section of Boston") and taken to malicious websites for theft of credentials. This is especially effective in mobile communication channels like WhatsApp, WeChat, SMS, or iMessage that warrant immediate responses from perceived authorities [Venkat, 2020].

#### 10.2.3.2 *Contact tracing scam*

Singaporeans have been targeted by contact tracing scams, with scam callers impersonating officials from the Ministry of Health (MOH) with the aim of identifying and monitoring people who have been in close contact with individuals infected with COVID-19. However, instead of pertaining to relevant details, the scammers go on to ask for financial details [Emen, 2020]. In the same vein, households have been approached by scammers through WhatsApp messages or directly visited to gain entry or request money for conducting 'government-endorsed' activities. In some incidents, scammers have pretended to visit for the purposes of contact tracing or sterilisation of homes, but instead stolen valuables after being allowed into victims' houses.

Opportunistic scammers have also taken advantage of Singapore's newly developed 'TraceTogether' application, which allows for faster contact tracing through Bluetooth-enabled detection of other users. Individuals have received recorded messages in English or Chinese,

with speakers claiming to be from MOH, and possessing an important message accessible by pressing '9' on their mobile phone. While the true motivation underlying these fake calls is still unknown, apart from the spread of misinformation, it is possible that users who fall victim to the scam may unintentionally compromise access to their phone data, including their personal or financial information.

With strict legislation and laws in place (allowing for the imposition of fines, imprisonment, revocation of citizenship rights and deportation in extreme cases) for failure to provide accurate or complete information, individuals may abide by phony law enforcement officers to avoid punishment. In fact, as per the persuasion principle of consistency [Cialdini, 1984], whereby people tend to act in ways that are consistent with their past actions, individuals who have already provided details of their movements may simply continue to give financial details regardless of any suspicions they may harbour towards the line of questioning. Nevertheless, the fear of punishment may not be the only motivating factor leading to increased vulnerability to scam.

## 10.3 Our Altruistic Acts May Be Abused by Others

Secondly, our charitable spirit and desire to act altruistically to help others in this trying time may potentially be exploited by scammers. Altruism refers to one's ability to rise above the basal instinct of self-preservation and instead, care for others. Humanity has been challenged to rise above the chaos to make it through together through acts of altruism. To elaborate, this philanthropic capability is exhibited through altruistic motivated acts—a desire to do good for others even at the expense of oneself [Cepelewicz, 2016]. Given the difficult COVID-19 situation, the world has witnessed the community coming together to help each other tide through the pandemic, sometimes at the risk of their own lives. Individual and communal efforts to help each other have arisen in Singapore. An example of individual effort has been seen through some homeowners, like Angela Chan, who have opened their homes to Malaysians stranded in Singapore because of Malaysia's Movement Control Order on 18 March 2020 [Kiew & Bock, 2020] It is a selfless act that comes at the cost of personal privacy and perceived sense of security by letting strangers into their homes and lives. Meanwhile, an example of

communal effort would be Contribute.sg. It is a ground up initiative that galvanises the community to contribute free surgical masks and hand sanitisers, in a time of their scarcity, to the less fortunate [https://www.contribute.sg/, 2020].

In social exchange theory, it is hypothesised that social interaction between people is based on a cost-benefit analysis ["What is social exchange theory?," 2018]. Altruistic acts are also guided by this theory as we see in reciprocal altruism (performing good acts for others with the expectation of similar treatment in future). This is governed by reciprocal norms—social norms that remind us to adhere to reciprocal altruism [Jhangiani & Tarry, 2014].

It is human nature to focus on the similarities we share with others and choose to perform altruistic acts on both the macro (greater good of the human race) and micro (good will towards target individuals) levels. Dissecting human altruism into these two domains allows us to consider at depth how perpetrators exploit altruism in the face of the global COVID-19 outbreak.

### 10.3.1 *Micro-level altruism*

It is easy to think that the right thing to do is to always help another person in need. However, rewards and costs as well as narratives play important roles in determining whether altruistic acts are ultimately performed. Darley and Batson's 1973 results of their Good Samaritan experiment showed that individuals were less likely to perform acts of altruism if it was at a greater cost to themselves. Similarly, being able to identify with the person in need evokes reciprocal altruism as discussed above, benefitting the self in an evolutionary[a] way. While the definition of altruism is good done for others even at the expense of oneself, it is observed that we are still primed to choose the best possible option for ourselves within the perimeters of helping others.

[a]From an evolutionary perspective, altruism ensures the continuation of humankind, whereby helping those identified as kin (who share similarities with us) would encourage living to reproductive maturity to pass on genes to future generations, bringing about the survival fitness of the human race. This possible evolutionary role might be a reason why altruism exists in society. Altruism is social in nature and not limited to human societies alone; animals are also seen to participate in reciprocal altruism.

#### 10.3.1.1 *Personal donation scams*

Exploiting the goodwill of people by means of setting up fake personal donations for "survivors" is not unique to COVID-19 [Government of United Kingdom, 2020]. In studying the patterns of donations, people are more generous around festive seasons [Gordy, n.d.] where ask campaigns use the narrative of joy juxtaposed against the alluded sadness of the less fortunate, making the contrast poignant to viewers. It has been shown that narratives identifying the exact beneficiary is more important than facts of a need in persuading individuals to donate [Small, 2003], a point employed by scammers who create entire profiles (some with pictures) to gain sympathetic responses.

Scam donations for COVID-19 "survivors" (and "patients") anchor on relaying a narrative that highlights the "survivors" plight, allowing potential scam victims to form an emotional connection with the story. As liking and familiarity increases, potential victims are more likely to accede to requests for help [Cialdini, 1984] and donate to those "in need." This is particularly aided by social media or instant messaging platforms as vehicles to speak directly to victims without coming across as confrontational as when done in person, which increases chances of donations [Gordy, n.d.].

#### 10.3.1.2 *E-Commerce scams* (*revisited*)

On top of exploiting one's fear, e-commerce scams leverage on the persuasion tactic of scarcity, as mentioned previously. It is a safe assumption then that masks are something many people would lack, and if made available, those near and dear to them would attempt to purchase. This propensity to protect not just oneself but also in one's social circle, especially immediate dependants, leads to the exploitation of people's goodwill for their loved ones. By pricing items reasonably lower than competitors, scammers appeal further to the cost-benefit analysis that takes place in decision making that encourages buyers to choose them over costlier alternatives. According to Marshallian Economics, the lower cost also encourages buyers to buy more than initially wanted or needed, which translates to an unexpected ability to afford more, and fuelling goodwill to buy more masks for loved ones. This explains why some victims purchasing on behalf of their companies and loved ones were among those who lost the most money, and how some victims go on to introduce such fraudulent sellers to others [Chia, 2020].

### 10.3.2 *Macro-level altruism*

On a macro level, there are scams that run on the altruistic intent people hold for their wider community and even beyond the borders. In line with the findings of Lowery *et al.* [2006] that people tend to be more concerned with what happens to the group they belong to than an opposing group, we can observe how groups tend to do things that benefit their own group members.

#### 10.3.2.1 *Public donation scams*

In view of COVID-19, hoax CDC bitcoin donation campaigns in the United States have been set up to scam individuals of bitcoin funds [Vergelis, 2020]. Criminals behind this scam prey on the collective group identity of individuals who believe themselves to be a part of the community the CDC serves, as they are willing to do altruistic acts to seek or maintain positive social identity [Tajfel & Turner, 2004]. The desire to donate for the greater good of those who benefit from CDC's work can be seen as an altruistic act of social responsibility, the need to do their part to maintain their social identity. Scammers appeal to this strain of reciprocal altruism that is evoked from an identity narrative where individuals define themselves and are defined by others as members of the group enforcing reciprocal norms. Additionally, victims may be persuaded by the use of social proof[b] [Cialdini, 1984] as they are told of other community member donations with two consequences: firstly, to make the donation campaign seem credible, and secondly, to reinforce group norms (i.e., that everyone is donating) and validate their actions (i.e., they should donate too).

#### 10.3.2.2 *Contact tracing scam* (*revisited*)

Contact tracing scam is yet another scam that exploits human's needs for acts of altruism. Victims who believe that this is the same beneficial government initiative they have heard about would cooperate to help contain the spread of COVID-19 by disclosing their personal details openly [Personal Data Protection Commission Singapore, 2020]. Impersonation scams run by individuals fronting as government authorities

[b] Social proof is the assumption made when we are unsure of how to act that others around us have more knowledge on what should be done (i.e., when we are unsure of the correct behaviour and look to the behaviour of others as the standard for what is right/appropriate.)

target those who desire to do good for the community. Potential victims can fall into one of three categories: (1) highly empathetic individuals; (2) people who reason that complying is the right thing to do; and (3) those who took up the opportunity to act altruistically given the unique circumstances [Walker & Frimer, 2007]. This posits that even if a person were more emotional, more logical, or more opportunistic, their altruism could be taken advantage of by such impersonation scams.

## 10.4 Our Desire for More May Cost Us Even More

Finally, among all the responses to COVID-19, we see greed not only from scammers, who see this outbreak as an opportunity to cash in, but from victims of scams as well. Greed has been hypothesised to be a response to neglect, abuse, or lack of stability in early life and manifests later when anxiety is coupled with low self-esteem [Burton, 2014]. It is a fixation on a substitute that is no longer necessary and taken to excess. As long as there is a fear of lack, there will exist the possibility of greed in trying to accumulate enough to dispel those fears. An example is the panic buying in Singapore on 6 February 2020 where greediness exemplified in hoarding illustrates the fear of the lack of necessities. The government had to reassure the nation of its reserves to dispel fear which seemed to work. However, when news of Malaysia's Movement Control Order on 16 March, a smaller scale of the earlier panic buying ensued causing the government to step in, again, to reassure that Malaysia will continue exporting food to Singapore ["COVID-19: There's no need," 2020].

### 10.4.1 *Investment fraud*

Riding on the wave of COVID-19, scammers have claimed certain companies are close to finding the cure or vaccine to the virus in order to entice victims to make investments in them by purchasing their stocks and benefiting from future profits [Homeland Security Today, 2020]. Such scams are known as 'pump-and-dump' schemes, where stocks are artificially inflated to give the illusion of a company with profitable stocks sold at prices that are growing healthily. Economies and industries have been affected by COVID-19, resulting in anxiety about job security and a loss in some people's self-esteem at a time when it is tethered to the ability

to provide for themselves and/or their family. Citing Burton [2014], the mixture of the two can arouse greed within individuals who then look to alternatives in securing their self-esteem needs and eradicating anxiety. The plausibility of such companies existing backed with 'evidence' of healthy stock prices are designed to tempt individuals to act in greed, investing a sum of money that is small in comparison to the projected future earnings. Scammers employ the authority persuasion tactic [Cialdini, 1984] to gain their victim's confidence to trust the fraudulent investment presented by them parading as a professional with insider knowledge and experience.

## 10.5 Recommendations

With the fallout of COVID-19 pandemic expected to last for a long time, cybercrimes and scams might continue to permeate the virtual ecosystem and prey on the uninformed and anxious who turn to online sources for information and help without adequate knowledge and discernment. It would thus be vital to conceive of and implement measures that can combat the growing spate of such duplicitous acts, whether perpetrated by local and/or international operators. As illustrated above, these implementations might benefit from accounting for the individuals' psychological vulnerabilities, to allay such compromising tendencies and create safeguards for the community as a whole.

### 10.5.1 *Leverage on reliable and recognisable information-providing channels*

To combat fear and uncertainty towards the physical and cyber dimensions of the COVID-19 situation, efforts must be made to draw public attention and traffic to information platforms that are trusted, recognisable, and easily accessible to all. These sources could further consolidate credible and critical information from multiple sources on cyber-crimes and scams related to the COVID-19 situation, e.g., Singapore's ScamAlert website which has an established reputation on publishing news and experiences of scam encounters in the country. Additionally, popular platforms such as Channel News Asia and Gov.sg that are already well-visited by the community for their timely news and alerts on COVID-19 can further include an up-to-date section on

COVID-19 cybercrimes to leverage on the existing traffic. Fundraising platforms could similarly feature scam-related advisories and run stringent checks and verifications to ensure information and donation requests are legitimate. This would not only help to raise awareness, but also provide important advice, direct users to appropriate channels for enquiry or help and prevent individuals from being scammed.

Apart from delivering information over public and conventional online and print channels, organisations can also be roped in to play an important complementary role in raising and reiterating important messages on COVID-19—related cyber threats to their employees. The WHO Information Network for Epidemics (EPI-WIN) believes that working closely with large-scale multinational corporations to disseminate credible information is important as employees tend to view their employers as a trusted institution [Gharib, 2020]. Therefore, employers can reinforce key knowledge and combat misinformation among their employees. Concerning cybersecurity policies, they can further keep their staff informed on the nature of detected cybersecurity breaches to increase their vigilance, while sharing the measures they would or would not take during the outbreak, such as soliciting any personal information in external domains or interventions that bear semblance to common COVID-19 scams. These stipulations will help to create clarity and clear the air among employees who might receive unusual instructions or requests from spoofed business accounts or other impersonated sources.

### 10.5.2 *Employ a diversity of public messaging formats*

The efficacy of public communication messages can be increased by providing information in new or diverse ways to boost uptake and retention rates. Through tailored recommendations, the public can be advised against immediately engaging with potentially harmful platforms, and instead slowing down to first stop and think about the possible consequences of any hasty actions. For example, linking the notion of maintaining good personal hygiene to reduce the risk of contracting COVID-19 with upholding cyber hygiene to better avoid cyber threats can be a meaningful paradigm for individuals to keep themselves safe both offline and online.

The use of complementary and overlapping terms in both the physical and virtual prevention of viruses is often used to instil the mindset of

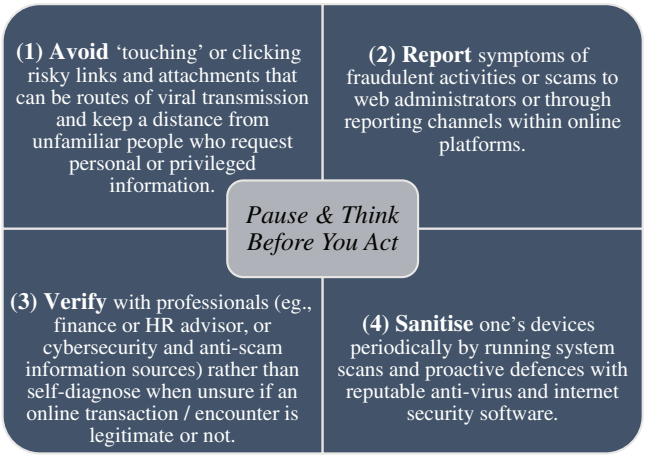



Figure 10.1. Recommendations on good cyber hygiene habits.

treating cyber hygiene behaviours as routine activities [Maennel *et al*., 2018]. Key messages can take the following form (see Figure 10.1).

The mode of communication can also matter when advocating messages that are aimed at the whole of society. Most international efforts to shine a spotlight on such COVID-19 related scams have been in simple, text-based formats, consisting of a brief description of the criminal modus operandi, sometimes supplemented by screenshots of fraudulent emails or messages. Usually, the notices also include tips to avoid falling victim, official contact details to report any scams encountered and any other relevant resources. While this approach may be adequate if the aim is to increase public awareness, it may not be the most suitable in engaging the audience as a preventive strategy. Delivering messages in creative formats can complement conventional text-based mediums as the added visual and dramatic elements may help to draw and sustain attention [Lu, 2015]. For example, the video on iconic local celebrity Phua Chu Kang dispensing advice on keeping oneself safe from the COVID-19 was well-received and widely circulated [Loh, 2020]. Therefore, strategically partnering with wide-reaching and trusted media channels in Singapore (e.g., Mothership) can achieve both the aims of disseminating key information to a larger audience as well as giving room for creative space to craft important messages that the public should take home. Adopting varied formats in public messaging campaigns may engage multiple

routes of information processing that can increase the salience and retention of information [Rice & Atkin, 2013].

### 10.5.3 *Proactive intervention by multiple parties*

Implementing upstream measures by government agencies and businesses continue to remain key in disrupting cybercrime activities leveraging on COVID-19 anxieties and providing additional safeguards to citizens. For example, the requirement for all Singapore telecommunication companies to add the plus '+' prefix to international calls is one upcoming measure to warn users on spoofed local phone numbers [Abdullah, 2020]. Tech companies can also continue to refine their systems and search terms algorithm to increase the prominence of credible information sources or pages and filter dubious or unverified sources. This could be important to better manage the influx of new domain names with terms (such as 'coronavirus') that have been registered, or new e-commerce posts related to the sale of an item or provision of service linked to COVID-19. Companies can also conduct proactive investigations to detect emails with content in the subject line or main body text that are commonly used by cyber criminals, such as the combination of the terms 'password change' and 'coronavirus' and an external link within the same email [Rayome, 2019].

## 10.6 Conclusion

The emergence of COVID-19 variants of familiar and novel cybercrime and scams depicts the ever-evolving landscape of cyberspace where criminals continue to innovate in their operations to prey on vulnerable systems and individuals. In Singapore, we are privileged to have many existing and upcoming initiatives, strategies, and policies from both the public and private sector to respond to these pervasive threats. These are multi-faceted and comprehensive approaches borne out of inter-ministry, inter-agency, and public-private collaborations and efforts. Therefore, the issue of mitigating cybercrimes and scams and safeguarding our citizens against them continues to remain a priority, particularly during these present times when the incidence of such threats may be increased.

As highlighted in this chapter, the human elements of fear, altruism, and greed may influence one's vulnerability to online threats during

this pandemic. It is hoped that raising awareness on these matters can help individuals to identify and understand how their emotions and predispositions may influence their decision-making and actions in the context of responding to a potential fraud. Therefore, it is without a doubt that individuals play a pivotal role in reducing and preventing victimisation to cybercrimes and scams by being self-aware and exercising due diligence when interacting with online sources, platforms, or users.

## 10.7 Acknowledgement

The views expressed in this chapter are the author's only and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 10.8 References

4 people arrested over Carousell face mask scams. (2020, March 7). *CNA.* https://www.channelnewsasia.com/news/singapore/coronavirus-covid-19-carousell-face-mask-scam-police-12512150

Abdullah, Z. (2020, March 3). International calls to have plus sign prefix to combat scam calls: Janil Puthucheary. *CNA*. https://www.channelnewsasia.com/news/singapore/international-calls-plus-prefix-combat-scams-janil-puthucheary-12494734

Ang, P. (2020, March 15). Two teenagers arrested for cheating victims of more than $9,800 in face mask and USS tickets scams. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/two-teenagers-arrested-for-cheating-victims-of-more-than-9800-in-face-mask

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), pp. 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Burton, N. (2014, October 6). Is greed good?: The psychology and philosophy of greed. *Psychology Today*. https://www.psychologytoday.com/sg/blog/hide-and-seek/201410/is-greed-good

Buxton, D. (2020, February 3). Coronavirus used to spread malware online. *Kaspersky*. https://usa.kaspersky.com/blog/coronavirus-used-to-spread-malware-online/20213/

Cepelewicz, J. (2016, March 3). What's your real motive for being altruistic? *Scientific American.* https://www.scientificamerican.com/article/what-s-your-real-motive-for-being-altruistic/

Chan, M., Regalado, F., & Cheng, T. (2020, March 4). Coronavirus scams prey on the fearful in China, Japan and beyond. *Nikkei Asian Review.* https://asia.nikkei.com/Spotlight/Coronavirus/Coronavirus-scams-prey-on-the-fearful-in-China-Japan-and-beyond

Chia, O. (2020, March 12). Coronavirus: Buyers of cheap masks lose $122 000 to overseas scam. *The Straits Times.* https://www.straitstimes.com/singapore/coronavirus-buyers-of-cheap-masks-lose-122000-to-overseas-scam

China has seized 31 million fake face masks amid crisis. (2020, February 27). *The New Paper.* https://www.tnp.sg/news/world/china-has-seized-31-million-fake-face-masks-amid-crisis

Cialdini, R. B. (1984). *Influence: The Psychology of Persuasion.* HarperCollins.

Clements, M. (2013). Self-Interest vs. Greed and the Limitations of the Invisible Hand. *American Journal of Economics and Sociology*, *72*(4), pp. 949–965. https://www.jstor.org/stable/23526068

*COVID-19: There's no need to rush to buy essentials*. (2020, March 17). Gov.sg. https://www.gov.sg/article/covid-19-theres-no-need-to-rush-to-buy-essential-items

Darley, J. M., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behaviour. *Princeton University Journal of Personality and Social Psychology, 27*(1), pp. 100–108. https://greatergood.berkeley.edu/images/uploads/Darley-JersualemJericho.pdf

Emen, M. (2020, February 16). Don't Fall Prey to These 5 Cruel Coronavirus Scams.*CCN.com*.https://www.ccn.com/dont-fall-prey-to-these-5-cruel-coronavirus-scams/

Gandel, S. (2020, January 29). Facebook struggles to stem spread of coronavirus misinformation. *CBS News*. https://www.cbsnews.com/news/facebook-coronavirus-posts-spread-misinformation-on-deadly-outbreak/

Gatlan, S. (2020, March 19). WHO chief impersonated in Phishing to deliver Hawkeye malware. *BleepingComputer*. https://www.bleepingcomputer.com/news/security/who-chief-impersonated-in-phishing-to-deliver-hawkeye-malware/

Gharib, M. (2020, February 21). Fake Facts Are Flying About Coronavirus. Now There's A Plan To Debunk Them. *NPR.org.* https://www.npr.org/sections/goatsandsoda/2020/02/21/805287609/theres-a-flood-of-fake-news-about-coronavirus-and-a-plan-to-stop-i

Gordy, Jeff. *10 year-end giving statistics every fundraiser should know*. Neoncrm.com. https://www.neoncrm.com/10-year-end-giving-statistics-every-fundraiser-should-know/

Government of Switzerland. World Health Organization. (2020a, April 26). Coronavirus disease 2019 (COVID-19) Situation Report—108. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200507covid-19-sitrep-108.pdf?sfvrsn=44cc8ed8_2

Government of Switzerland. World Health Organization. (2020b). Cybersecurity: Beware of criminals pretending to be WHO. https://www.who.int/about/communications/cyber-security

Government of United Kingdom. Bournemouth, Christchurch and Poole Council. (2020). Scams Coronavirus (COVID-19). https://www.bcpcouncil.gov.uk/News/News-Features/COVID19/Scams-to-avoid-Coronavirus-COVID-19.aspx

Holmes, A. (2020, March 9). Email scammers are taking advantage of coronavirus fears to impersonate health officials and trick people into giving up personal information. *Business Insider Singapore.* https://www.businessinsider.sg/coronavirus-email-scam-covid-19-phishing-false-information-who-cdc-2020-2?r=US&IR=T

Homeland Security Today. (2020, February 25). *Scammers Take Advantage of COVID-19 Outbreak to Carry Out Fraud*. https://www.hstoday.us/channels/global/scammers-take-advantage-of-covid-19-outbreak-to-carry-out-fraud/

ICE71. (2020, February 18). *Phishing scams around COVID-19*. https://ice71.sg/phishing-scams-around-covid-19/

Izard, C. E. (1991). *The psychology of emotions.* Springer Science & Business Media.

Jhangiani, R., & Tarry, H. (2014, August). Understanding altruism: Self and other concerns. *Principles of Social Psychology—1st International Edition*. https://opentextbc.ca/socialpsychology/chapter/understanding-altruism-self-and-other-concerns/

Kiew, C., & Bock, H. (2020, March 25). The life list: 7 kind acts by Singaporeans during the COVID-19 outbreak. *The Straits Times.* https://www.straitstimes.com/lifestyle/entertainment/the-life-list-7-kind-acts-by-singaporeans-during-the-covid-19-outbreak

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), pp. 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Loh, K. K. (2020, February 16). Coronavirus: TV characters Phua Chu Kang and Liang Ximei rally people to do right thing. *The Straits Times*. https://www.straitstimes.com/lifestyle/entertainment/coronavirus-tv-characters-phua-chu-kang-and-liang-ximei-rally-people-to-do

Lowery, B. S., Unzueta, M. M., Knowles, E. D., & Goff, P. A. (2006). Concern for the in-group and opposition to affirmative action. *Journal of Personality and Social Psychology*, *90*(6), pp. 961–974. https://pdfs.semanticscholar.org/292a/55699566ee2815492a2acc5deb973fc73951.pdf?_ga=2.123764105.2056493896.1583979787-1013451908.1583979787

Lu, S. (2015). An epidemic of fear. *Monitor on Psychology*, *46*(3), p. 46. https://www.apa.org/monitor/2015/03/

Maennel K., Mäses S., & Maennel O. (2018). Cyber hygiene: The big picture. In Gruschka N. (Ed.), *Secure IT Systems. NordSec 2018. Lecture Notes*

*in Computer Science, vol 11252*. Springer. https://doi.org/10.1007/978-3-030-03638-6_18

McDaniels, T., Axelrod, L., Cavanagh, N., & Slovic, P. (1998). Perception of ecological risk to water environments. *Insurance: Mathematics and Economics*, *22*(2), pp. 190–191. https://doi.org/10.1016/s0167-6687(98)80048-6

Personal Data Protection Commission Singapore. (2020, February 13). *Advisory on Collection of Personal Data for COVID-19 Contact Tracing*. https://www.pdpc.gov.sg/Advisory-on-CUD-for-COVID-19

Rayome, A. D. (2019, March 19). How to prevent spear phishing attacks: 8 tips for your business. *TechRepublic*. https://www.techrepublic.com/article/how-to-prevent-spear-phishing-attacks-8-tips-for-your-business/

Reeves, C. (2020, March 17). Watch out for Coronavirus (COVID-19) phishing and malware. *Telstra Exchange*. https://exchange.telstra.com.au/watch-out-for-coronavirus-covid-19-phishing-and-malware/

Rice, R. E., & Atkin, C. K. (2013). *Public communication campaigns* (4th ed.). Sage Publications. http://dx.doi.org/10.4135/9781544308449

Rollert, J. P. (2014, April 7). Greed Is Good: A 300-Year History of a Dangerous Idea. *The Atlantic*. https://www.theatlantic.com/business/archive/2014/04/greed-is-good-a-300-year-history-of-a-dangerous-idea/360265/

Rotfeld, H. J. (1988). Fear appeals and persuasion: Assumptions and errors in advertising research. *Current issues and research in advertising*, *11*(1–2), pp. 21–40.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), pp. 1–28. https://doi.org/10.1037/h0092976

Sarraf, S. (2020, January 14). Cyber scammers take advantage of bushfire crisis. *CIO*. https://www.cio.com/article/3514051/cyber-scammers-take-advantage-of-bushfire-crisis.html

Schroder, H. M., & Rotter, J. B. (1954). Generalization of expectancy changes as a function of the nature of reinforcement. *Journal of Experimental Psychology*, *48*(5), pp. 343–348. https://doi.org/10.1037/h0058823

Schwartz, M. S. (2020, March 11). Missouri Sues Televangelist Jim Bakker For Selling Fake Coronavirus Cure. *NPR.org.* https://www.npr.org/2020/03/11/814550474/missouri-sues-televangelist-jim-bakker-for-selling-fake-coronavirus-cure?t=1584337157917

Seuntjens, T. G., Zeelenberg, M., Breugelmans, S. M., & van de Ven, N. (2014, September 4). Defining greed. *British Journal of Psychology*, The British Psychological Society. https://www.academia.edu/14429789/Defining_greed?-auto=download

Somesh. (2020, February 25). Coronavirus: Investment fraud and cyber scam warnings. *International Chamber of Commerce*. https://icc-ccs.org/index.php/1288-coronavirus-investment-fraud-and-cyber-scam-warnings

Slovic, P. (1987). Perception of risk. *Science*, *236*(4799), pp. 280–285. https://doi.org/10.1126/science.3563507

Slovic, P. (2010). *The Feeling of Risk: New Perspectives on Risk Perception*. London, Earthscan.

Small, D. A., & Loewenstein, G. (2003, January). Helping *a* Victim or Helping *the* Victim: Altruism and Identifiability. *Journal of Risk and Uncertainty*, *26*, pp. 5–16. https://doi.org/10.1023/A:1022299422219

Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In J. T. Jost & J. Sidanius (Eds.), *Key readings in social psychology. Political psychology: Key readings*, pp. 276–293. Psychology Press. https://doi.org/10.4324/9780203505984-16

Tham, I. (2020, February 1). Wuhan virus: Hackers exploiting fear of bug to target computers, gadgets. *The Straits Times.* https://www.straitstimes.com/tech/wuhan-virus-hackers-exploiting-fear-of-bug-to-target-computers-gadgets

Tidy, J. (2020, March 13). Coronavirus: How hackers are preying on fears of Covid-19. *BBC News.* https://www.bbc.com/news/technology-51838468

U.S. Securities and Exchange Commission. (2020, February 25). *Look Out for Coronavirus-Related Investment Scams—Investor Alert*. https://www.sec.gov/oiea/investor-alerts-and-bulletins/ia_coronavirus

Venkat, A. (2020, February 20). Phishing Campaigns Tied to Coronavirus Persist. *BankInfoSecurity*. https://www.bankinfosecurity.com/phishing-campaigns-tied-to-coronavirus-persist-a-13741

Vergelis, M. (2020, February 7). Coronavirus phishing. *Kaspersky*. https://www.kaspersky.com/blog/coronavirus-phishing/32395/

Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, *93*(5), pp. 845–860. https://doi.org/10.1037/0022-3514.93.5.845

Weisbaum, H. (2020, February 19). How to avoid falling victim to a coronavirus email scam. *NBC News.* https://www.nbcnews.com/better/lifestyle/how-avoid-falling-victim-coronavirus-phishing-email-attack-ncna1137941

What is social exchange theory? (2018, April 20). *Tulane University School of Social Work.* https://socialwork.tulane.edu/blog/social-exchange-theory

Wong, P. T. (2020, February 15). Faceless 'Santa Clauses' make grocery runs for family observing home quarantine. *Today Online.* https://www.todayonline.com/singapore/faceless-santa-clauses-make-grocery-runs-family-observing-home-quarantine

Zhang, J. (2020, February 19). SPF received more than 90 reports on alleged scam website MedicalLex, is looking into it. *Mothership.sg*. https://mothership.sg/2020/02/medicallex-scam-update/

This page intentionally left blank

# Part 3

# Assessment, Intervention, and Prevention

This page intentionally left blank

# Section E

# Insights for Assessment, Prevention, and Intervention

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0011

# Chapter 11

# Legal Issues and Ethical Considerations in Cyber Forensic Psychology

Benjamin Ang

*Centre of Excellence for National Security, S. Rajaratnam*
*School of International Studies, Nanyang Technological University*

*isbang@ntu.edu.sg*

## 11.1 Introduction

### 11.1.1 *The case of Gary McKinnon*

Between 2001 and 2002, Gary McKinnon scanned thousands of US government machines including PCs within the Army, Navy, Air Force, NASA, and the Department of Defense, copying files and passwords, and (at one point) shutting down the US Army's entire Washington, DC, network, for three days. McKinnon was traced to his apartment in London, where he was arrested by the United Kingdom's National Hi-Tech Crime Unit. They would then have extradited him to the United States for trial, except that he was diagnosed with Asperger's syndrome, a form of autism, as well as depression. He had apparently intruded on those computer systems so that he could find and release hidden government information on alien antigravity devices and advanced UFO energy technologies, for the benefit of humanity [Kushner, 2011].

National leaders were divided on the role of Asperger's in McKinnon's crime, and whether he deserved sympathy or a strong penalty [Kushner, 2011]. Finally, in 2012, the UK home secretary announced that McKinnon would not be extradited on human rights grounds, because medical reports had warned that he was at risk of suicide if sent to face trial in the United States, and the UK Crown Prosecution Service announced no further legal action would be taken in Britain against him because of lack of evidence in the UK. McKinnon still cannot travel outside of the UK because the extradition warrant is still outstanding [Kennedy, 2012].

McKinnon's case illustrates the connection between cyber deviant behaviour (he was the perpetrator of a cybercrime) and forensic psychology (the diagnosis of Asperger's and depression was relevant to his ability to stand trial in the United States). Experts believe that people with Asperger's "might be quite skilled at hacking" because they love systems, including computer systems, and many of them have an obsessive interest in technology, physics, and space [Kushner, 2011]. Therefore, it may be useful to apply forensic psychology in cybercrime prevention, detection, and response. In particular, this chapter explores the legal and ethical considerations of this application. Since forensic psychology can be further supported and enhanced by cyber tools, this chapter will also examine the legal and ethical considerations of doing so.

## 11.2 Scope and Definitions

For the purposes of this discussion, we will use certain definitions of terms:

- *Cyber psychology* is "the study of human interactions with the Internet, mobile computing and telephony, games, virtual reality, artificial intelligence, and other electronic technologies … it assesses how we interact with others using technology, how our behaviour is influenced by technology and how our psychological states can be affected by technologies" [Power & Kirwan, 2014, p. 3].
- *Forensic psychology* is psychological knowledge which can be applied in the criminal justice system to the prevention of crimes, investigation, legal decision-making, and rehabilitation of convicts [International Journal of Psychology, 2016].

- *Investigative psychology* and *offender profiling* are aspects of forensic psychology used to infer characteristics of a criminal based on his or her behaviour during the crime [Winerman, 2004].

## 11.3 Using Forensic Psychology to Respond to Cybercrime

Forensic psychology has been useful in investigating and prosecuting crimes in general. When applied to cybercrime, it could help to determine why an individual may choose to engage in cybercrime, devise potential methods for deterrence, and predict effects of crimes on victims [Power & Kirwan, 2014].

For example, as part of an investigation, forensic psychology could build up the profile of a cybercriminal, which would include key elements such as personality traits that predispose him/her to committing cybercrime, personality traits that enable him/her to commit cybercrime effectively, motivation, social characteristics, demographic features, socioeconomic status, and social or moral qualities [Kipane, 2019].

However, some experts warn that profiling of cybercriminals is risky. When studying other types of crimes, profiling assumes that (1) human behaviour is constant; and (2) specific clues to the offender's psychology can be obtained at the crime scene [Kipane, 2019]. In cybercrime cases, the software and hardware tools used by cybercriminals can be configured to automatically randomise behaviour (such as time of day of attack, or duration of attack), and even to leave false clues at the 'scene of the crime' about the offender's identity or characteristics (such as faking the country of origin, or language used). Investigators have an ethical obligation to recognise that there is some factual uncertainty attached to attribution.

With that caveat in mind, we will examine how cyber forensic psychology can be applied to three main types of cybercrime: (1) crimes that exist in the offline world (fraud, child pornography, harassment), but are now enhanced by the use of technology; (2) crimes against computer systems (hacking, malware development) [Power & Kirwan, 2014]; and (3) new crimes that are enabled by technology (cyberbullying).

### 11.3.1 *Different aspects of cybercrime*

(1) *Cybercrime includes crimes that exist in the offline world* (*fraud, child pornography, harassment*)*, but are now enhanced by the use of technology.* Most criminal law systems, such as Singapore's Penal Code, recognise crimes such as fraud, child pornography, and harassment as offences, without distinguishing between whether they are committed online or in person.

(2) *Cybercrime includes crimes committed against computer systems.* Many jurisdictions now recognise crimes committed against computer systems as offences. For example, the Computer Misuse Act [Chapter 50A, Statutes of the Republic of Singapore, Revised Edition 2007] (hereafter "CMA") specifically covers crimes against computer systems, and "for securing computer material against unauthorised access or modification and for matters related thereto." For ease of reference, we will refer to offenders in this category as 'hackers.' The main offences under the CMA are listed in Table 11.1.

(3) *Cybercrime includes new offences that have been enabled by technology.* New deviant behaviours such as cyber stalking or

Table 11.1. Main offences under the Computer Misuse Act (CMA).

| Section | Offence |
|---|---|
| Section 3 | Unauthorised access to computer material |
| Section 4 | Access with intent to commit or facilitate commission of offence (such as fraud or theft) |
| Section 5 | Unauthorised modification of computer material (including deleting or changing data) |
| Section 6 | Unauthorised use or interception of computer service (including using networks without permission) |
| Section 7 | Unauthorised obstruction of use of computer (including distributed denial of service attacks) |
| Section 8 | Unauthorised disclosure of access code (including selling of stolen passwords and PINs) |
| Section 8A | Supplying, etc., personal information obtained in contravention of certain provisions (including dealing with stolen personal information) |
| Section 8B | Obtaining, etc., items for use in certain offences (including making or selling of malware or hacking tools) |

harassment, and 'doxxing' (publishing information about a person's identity to cause fear or provoke violence against the person) [Neo, n.d.], have been enabled by the widespread use of digital technology, social media platforms, and messaging tools. Many jurisdictions have passed new legislation to criminalise these behaviours, such as Singapore with its Protection from Harassment Act (Chapter 256A). In addition, some jurisdictions (Singapore again) are seeking to amend their laws to deal with "emerging crime trends" involving technology, such as voyeuristic photo/video taking (upskirt), recording of people in intimate moments without consent, the distribution of intimate photos/videos without a person's consent, and fraud arising from sophisticated deceptive schemes where wrongful gain or loss is intended, without an identifiable victim being deceived [Elangovan, n.d.].

### 11.3.2 *Review of psychological research on cybercrimes*

To date, researchers have not identified individual personality traits which are reliable indicators of potential online fraudsters. While common elements could include financial strain, ego or power, and neutralisation (justifying actions by suggesting that the victim can afford to be or deserves to be defrauded), these are insufficient to assist in the detection or prevention of fraud. The psychology of fraud victims is also unclear—while some findings are discussed below, more research is needed to determine what psychological factors make some people more likely to fall prey to fraudsters, while others less likely [Kirwan & Power, 2014].

On the other hand, more extensive research has been done on the profile of child pornography offenders. These offenders may have cognitive distortions whereby they justify the offending behaviour to themselves. There are also studies of their demographics and psychometric traits, all of which can be used in profiling offenders as well as in rehabilitation efforts. Nevertheless, more research is still needed to determine which rehabilitation methods help to reduce recidivism, as well as to understand the psychological impact on victims (the subjects of the photos/videos, whose images are now circulating online), and how to treat them [Kirwan & Power, 2014].

In terms of hacking, some researchers have proposed that understanding the psychology of hackers can help to improve security measures. However, research on hackers is difficult because offenders

who have been caught may not respond to questions truthfully; offenders who have not been caught do not wish to disclose information which could lead to their capture; and some online respondents may be people who pose as hackers (for various reasons) but are not actually hackers. As a result, there is still relatively little empirical evidence about hackers; what exists is contradictory, and it is difficult to develop an overall picture of what a hacker is like. Many of the theories need to be empirically investigated before they can be considered reliable [Kirwan & Power, 2014]. The amount of research on people who develop malware is even less than the research on hackers [Kirwan & Power, 2014].

In the area of cyberbullying or harassment, there has been considerable research into the psychology of offenders—examining personality traits, cognitive abilities, and motives—and also into the psychological effects on victims. This research can be used to help develop preventative measures and responses, such as counselling for victims and treatment for offenders. As many of these efforts have focused on children and young people so far, more focused research is needed on cyberbullying or harassment in the workplace [Kirwan & Power, 2014], which is another fertile ground for this problem.

As further research continues, researchers need to keep in mind three main ethical issues: (1) informed consent (especially if the subjects are persons involved in illegal or deviant behaviours which could lead to arrest); (2) impact of the research design on outcome, especially in random assignment of human subjects (whether offenders or victims); and (3) guarantees of confidentiality and immunity (which tend to improve the quality of responses, but also prevent the researchers from reporting crimes) [Vohryzek-Bolden, 1997].

## 11.4 Using Cyber Tools to Enhance Forensic Psychology

One of the benefits of the research described above is the enhancement of criminal profiling. Criminologists and law enforcement officers use criminal profiling, which is based on the analysis of behavioural and psychological patterns, to predict the characteristics of a suspect. This analysis is used to help identify criminal behaviour or actions, and to determine the criminal's personality, his/her modus operandi, and possible motivations.

Police and academics differ in opinion on whether these inferences should come from peer-reviewed research or from investigative experience [Winerman, 2004]. Cyber-enabled forensic psychology could offer an alternative—collecting data from known offenders' social media posts, mobile phone metadata, messages, and other data, to build a large dataset for analysis.

Corporations (e.g., Amazon, Facebook, or Google) already collect and mobilise Big Data to predict future human behaviour through systems based on artificial intelligence and machine learning, in order to target advertising and influence purchasing [Gstrein *et al*., 2019]. If this technology is harnessed for cyber forensic psychology, it could potentially be used to predict crime, identify criminals and victims, target messaging, and influence behaviour. However, in some jurisdictions, this use of data could conflict with data privacy laws or laws regulating search and surveillance. It could also raise ethical issues of using users' data for purposes they had not consented to.

### 11.4.1 *Legal and ethical issues in data collection*

Collection of data can be difficult in jurisdictions like the US, where the Fourth Amendment to their Constitution prohibits "unreasonable searches, and seizures of property" by the government. The US courts have interpreted this in ways limiting the ability of law enforcement to collect historical cell site location information (CSLI) of customers held by mobile phone companies [Liptak, 2018], GPS tracker data [Meyer, 2015], or the contents of a mobile phone [Dixon, n.d.].

On the other hand, jurisdictions like Singapore have no such constitutional restrictions. Instead, under Section 39 of the Criminal Procedure Code (Chapter 68), "a police officer or an authorised person investigating an arrestable offence may, at any time—

(a) access, inspect and check the operation in or from Singapore of a computer (whether in Singapore or elsewhere) that the police officer or authorised person has reasonable cause to suspect is or has been used in connection with, or contains or contained evidence relating to, the arrestable offence; and
(b) use any such computer in or from Singapore, or cause any such computer to be used in or from Singapore—(i) to search any data contained in or available to such computer; and (ii) to make a copy of any such data."

Provided that the criteria are met (investigation of an arrestable offence, and reasonable cause to suspect the computer contains evidence), this gives the police wide-ranging powers to collect data from relevant computers, mobile devices, and other computer systems. Without this provision, the police would be contravening Section 3 of the Computer Misuse Act. This provision also overrides the Personal Data Protection Act (PDPA), which requires consent to be given for the collection of personal data. In any case, the PDPA does not apply to the actions of the police.

While this would enable police to study individual suspects' social media or mobile devices, it is unlikely to justify wider data scraping (e.g., of public Facebook groups), even if there is reasonable cause to suspect that such data scraping would help collect relevant evidence. To date, this legal issue has not been addressed by the Singapore courts.

For the purposes of academic research, while researchers have regularly relied on 'public data', there are still ethical considerations, as a person's act of publishing online does not necessarily mean that he/she gives permission to be included in research, since he/she may have some expectations of anonymity [Dennen, 2012].

### 11.4.2 *Data collection at crime scenes*

When computer systems or computer networks are the 'crime scene', their scene conditions (server logs, metadata) and other investigation data can provide information about the personality, motivation, and characteristics of the offender (provided the cybercriminal has not used tools to plant false and misleading clues). A multidisciplinary team of specialists—including psychologists, technology specialists, and the police—is essential. The profiler team analyses the data to provide predictions of the nature of a potential criminal, and this process can either be "clinical" (using profilers' intuition, knowledge, experience, and training to generate predictions) or "statistical" (working with statistical data; databases on similar crimes) in nature [Kipane, 2019].

Profilers may also use digital behavioural analysis, a relatively new field that applies the concepts of traditional behavioural analysis to the digital footprints of criminals. Also known as Idiographic Digital Profiling (IDP), such analysis attempts to understand the behaviours of cybercriminals, examining a particular subject's digital footprints for immediate use in an ongoing investigation—i.e., subject identification,

lead generation, obtaining and executing warrants, and prosecuting offenders. In the case of 'Dread Pirate Roberts', investigators examined evidence which included the suspect's online postings, Google login data, and YouTube videos, to carry out cross-site tracking, identify his posts, map out the criminal enterprise, and enumerate his associates [Steel, 2014].

As mentioned earlier in this chapter, advanced cybercrime techniques include leaving false 'clues' to mislead investigators, such as launching the attack from unrelated servers in third-party countries, planting misleading foreign language texts in the source code of malware, or even posting fake messages and leaking fake documents on social media to confuse investigators [Greenberg, 2018]. Profilers in cybercrime cases have an ethical duty to recognise this possibility, and courts prosecuting cybercrime cases must also take this into consideration when weighing evidence.

## 11.5 Use of Forensic Psychology in Preventing Cybercrime

Regardless of whether a clinical, statistical, or digital behavioural approach is taken, forensic psychology could be used to prevent cybercrime. The well-known "Routine Activity Theory" of criminology proposes that a crime will take place whenever a motivated offender encounters a suitable target (victim) in the absence of guardians [Cohen & Felson, 1979].

### 11.5.1 *Psychology of the victim (the "suitable target" or the "guardian")*

Using online fraud as an example, victims can fall prey to online fraudsters if they are enticed by prospects of immediate gratification (monetary, sexual), under the influence of optimism bias, and assign lower risk values to privacy decisions [Wiederhold, 2014]. While some studies have shown that impulsivity may indicate susceptibility to phishing and online scams, there appears to be no strong connection between online scam victimisation and specific personality traits and impulsivity factors [Kirwan *et al*., 2018]; i.e., anyone can become a victim. In fact, those in positions of trust and authority, who are

well-educated and over-confident of their intellectual superiority, such as senior partners in a law firm, may be more vulnerable to the social engineering tricks of phishing and online scams [Alashe & Cross, 2019].

Experts instead suggest that psychologists can help by increasing our understanding of both victims and offenders, working with technology providers to develop systems, and with legislators to develop policies that can help to achieve better security [Wiederhold, 2014].

As we understand more about the psychology of scams, the need to set up 'guardians' for ourselves arises. For example, studies show that phishing emails are "peripherally" processed, i.e., we are focused on the perceived urgency of the message, rather than whether the message is authentic [Attrill-Smith *et al*., 2019]. This can happen in high-pressure work environments with high standards and a heavy workload, where quick decision-making is the norm. In one case, a company's accountant had received an email stating that she was going to get a phone call from a consultant working with a lawyer, who would then give her 'confidential instructions' from the company president about transferring up to €500,000. She received 10 emails and four phone calls in one hour, which pressured her into reacting quickly without thinking [Keyworth & Wall, 2016].

Corporate training on phishing and cyber scams has focused on awareness raising, but it may be overly optimistic to think that everyone would remember to follow cyber security processes even when under such internal or external pressures [Alashe & Cross, 2019].

Therefore, it may be necessary for the computer system to compensate by monitoring and recognising when the user appears to be conducting a cognitively demanding task, and encouraging extra caution for any critical decisions [Attrill-Smith *et al*., 2019]. In this example, the computer system would have had to identify that the accountant was under time pressure, and forced her to verify the authenticity of the money transfer request before allowing her to proceed. This type of friction can improve security, but may also be unpopular in organisations that value quick response.

Online 'romance scams' are another example where victims engage in emotional thinking instead of logical thinking. More specifically, victims are deceived into sending money to cybercriminals who offer them love. This is a multi-million-dollar business, where victims not only transfer huge sums of money to their online 'lovers', some even compromise themselves in the process, either by sending indecent photos, or by

embezzling money ["Woman Scammed of S$1.4m Arrested; Police Believe She Stole Money from Her Company," 2019].

Given that cybercriminals find their victims on social network platforms (like Facebook) and dating sites (like Tinder), those platforms have an ethical duty to protect their users. Studies indicate that victims are often highly educated, middle-aged, have traits of impulsivity and lack of self-control, and score high on sensation-seeking and addiction characteristics, compared to non-victims. Since dating sites and social media platforms already gather some of these data points for the purpose of targeted advertising. they could in turn use the data to identify vulnerable users, and make them aware of specific risks [Whitty, 2018]. For example, if a vulnerable user (40-year-old, well-educated, addicted to casual online games) starts receiving messages from a previously unknown person overseas, the system could warn the user that this could be a scam. Should those dating sites and social media platforms be unwilling to implement this for their customers, legislators may have to step in with new laws to compel them.

### 11.5.2 *Diversion of the "motivated offender"*

The UK National Crime Agency has identified several psychological features of young offenders who might start cybercriminal careers, including motivations of "completing the challenge, sense of accomplishment, [and] proving oneself to peers." They then recommend deterrence through "positive opportunities, role models, [and] mentors," encouraging them to invest their programming skills into positive/prosocial activities instead [Attrill-Smith *et al*., 2019]. These could include hackathons (competitions where they can collaborate on cyber projects) or bug bounty programmes (where they can be invited to and rewarded for uncovering security vulnerabilities in computer systems).

In the field of online radicalisation, Google's Jigsaw subsidiary launched the Redirect Method, which uses targeted advertising and video confronting online radicalisation and discouraging young people from joining terrorist groups like ISIS. The Redirect Method combined Google AdWords technology and curated YouTube video content, with insights into how terrorists exploit recruits based on their insecurities, prejudices, and fears, so that vulnerable people who searched for ISIS keywords were instead redirected to information that refuted ISIS's propaganda [Dishman, 2019]. This concept could potentially be applied to some

narrow parts of cybercrime—for example, people searching for hacking tools, or how to commit cybercrime, could be redirected to information on legitimate hackathons and bug bounty programmes instead. More research and experimentation are needed in this field.

Jigsaw has stated that their technology does not identify potential ISIS recruits for surveillance and future arrest, but instead aims to provide them with better information so that they do not become ISIS recruits [Greenberg, 2016]. Applying this to young potential cybercriminals, there would be ethical problems in pre-judging who is carrying out these searches for intellectual stimulation, and who is doing so for criminal gain. This would in turn lead to legal problems in justifying surveillance or even arrest, if no crime has been committed, which brings us next to the controversial issue of predictive policing.

### 11.5.3 *Predictive policing*

Predictive policing began with the application of mathematical methods and insights from research on the occurrence of earthquakes to crime data in order to 'forecast crime'. It is focused on forecasting places where property crime or violent crime takes place, as well as which persons might be involved in criminal activity. Today, Big Data is collected and processed with artificial intelligence and machine learning to create automated decision-making in the prediction of crime and criminals. Theoretically, it could be applied to predict cybercrime, or to advise on sentencing for convicted cybercriminals, in the same way that police in Richmond, Virginia, have used it to predict violent crime [Meijer & Wessels, 2019].

Unfortunately, if the data collected is incomplete, inaccurate, or irrelevant, since machine learning by design is not transparent to humans, pre-existing unfair assumptions can be encapsulated in algorithms, resulting in unfair targeting of minorities and ethnic or religious groups. For example, the non-profit organisation ProPublica revealed that a tool used by some US judges to predict recidivism of prison inmates was based on an algorithm that had an inherent bias against black people, falsely tagging black defendants as future criminals [Gstrein *et al*., 2019].

If psychological data could be processed as well, it could help make predictive policing more accurate, or it could make it even more inaccurate and biased, depending on the quality of the data collected.

The main ethical issues of data selection, machine bias, interpretation of forecasts, transparency, and accountability, must first be resolved. As long as there is uncertainty about the accuracy of the data or the objectivity of the process, any predictive decisions made for policing or sentencing will be open to legal challenge and appeal.

## 11.6 Role of Cyber Forensic Psychology Evidence at Trial and Sentencing

Finally, when the cybercriminal has been apprehended, cyber forensic psychology evidence could be relevant at the trial, and the relevant psychologist may be called to give expert evidence. In Singapore, under Section 47 of the Evidence Act (Chapter 97), experts may be called upon to give their opinion on a point of scientific, technical, or other specialised knowledge. The Act defines an expert as "a person with such scientific, technical or other specialised knowledge based on training, study or experience." We should be sceptical of broad claims of expertise in forensic cyber psychology, because it requires multi-disciplinary skills and experience [Kirwan, 2018], which may sometimes be found in a team instead of an individual.

From the case law, the expert has a duty to be independent and unbiased when forming his opinion, to consider all relevant facts and materials to the case, including those that may diminish his/her opinion, and to provide an opinion on matters that lie only within his/her expertise [Who is an Expert Witness and How to Use Expert Evidence in Singapore, 2019]. Experts must be non-partisan, impartial, and independent, not intentionally presenting false information, or selectively presenting information that favours the side that retained them [Forensic Experts Group, 2019].

In practice, there are also cyber forensic psychology "experts" who fall short in competency, training, and experience, who testify in areas outside the scope of their expertise. There are even "experts" who are willing to skew their application of scientific principles to support the client's case. Unfortunately, legal practitioners and judges with no relevant scientific qualifications or experience may find it difficult to evaluate and cross-examine complex scientific evidence presented [Forensic Experts Group, 2019]. Instead, they may rely on scientifically unrelated factors like "impressions of the analyst's demeanour and

credibility, like the ability to survive cross-examination," which has led to grave injustice in other fields [Lo, 2019].

With this in mind, Singapore has proposed amendments to the Criminal Procedure Code, which would introduce an accredited panel for expert psychiatrists testifying in criminal cases. Only psychiatrists who are members of the panel would be allowed to give expert evidence [Chen & Chua, 2018].

Returning to the story of Gary McKinnon, the autistic hacker; if an expert had persuaded a Singapore court that a subject was of 'unsound mind', then it would amount to a complete defence. Under Section 84 of the Penal Code (Chapter 224), "Nothing is an offence which is done by a person who, at the time of doing it, by reason of unsoundness of mind, is—(a) incapable of knowing the nature of the act; (b) incapable of knowing that what he is doing is wrong (whether wrong by the ordinary standards of reasonable and honest persons or wrong as contrary to law); or (c) completely deprived of any power to control his actions." In some cases, severe or moderate depression may not excuse a person from criminal liability, but is taken to impair a person's judgement and becomes a mitigating factor in sentencing [Lam, 2019].

Rehabilitation and restorative justice play an important part in crime prevention as well. In Singapore prisons, inmates with mental illnesses are treated by psychiatrists and medical officers while incarcerated, and given counselling and therapy by correctional rehabilitation specialists and psychologists. The Institute of Mental Health's Forensic Psychiatry Community Service provides continued treatment and service to ex-convicts with a psychotic, affective, or anxiety disorder diagnosis, supported by case managers and medical social workers. Courts can also impose mandatory treatment orders (MTOs) for up to three years, whereby offenders must undergo psychiatric treatment as an alternative to imprisonment [Lam, 2019].

## 11.7 Recommendations

This chapter has shown several examples of how forensic psychology can be used for prevention, detection, and response, with respect to cybercrimes like online fraud, hacking, and cyber bullying. As more research needs to be done to develop this field, researchers must keep in mind ethical issues such as informed consent, impact of research design, and guarantees of confidentiality and immunity. Researchers can benefit

from collecting large amounts of digital data through cyber-enabled forensics, but must comply with laws and ethical norms on privacy, search, and surveillance.

Investigators can use evidence from cybercrimes to apply behavioural analysis to cybercriminals, but they must recognise that some or all of the evidence could be doctored or faked. When seeking to adduce this evidence in court, they also have a legal and ethical duty to disclose this.

Behavioural analysis of victims of cybercrimes, like phishing, online fraud, and romance scams, can be used to develop systems to protect victims and prevent crime. If social media platforms and technology providers refuse to develop these systems, legislators may need to pass laws to compel them.

Authorities can use psychological data, incorporated into big data, machine learning, and artificial intelligence, to aid in decision-making for policing, investigating, and sentencing. However, they must address ethical issues of transparency, bias in data selection, and accountability.

In sum, it is clearly beneficial to apply forensic psychology in cybercrime prevention, detection, and response, as well as to enhance forensic psychology with cyber tools, provided that the ethical and legal issues listed in this chapter are considered.

## 11.8 References

Alashe, O., & Cross, T. (2019, July 31). Tackling the Human Aspect of Cyber Security: Legal Psychology. https://www.cybsafe.com/whitepapers/whitepaper-legal/

Attrill-Smith, A., Fullwood, C., Keep, M., Kuss, D., & Kirwan, G. (2019). Rise of Cybercrime. *The Oxford Handbook of Cyberpsychology*. Oxford University Press.

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, *44*(4), pp. 588–608. http://doi.org/10.2307/2094589

Dennen, V. P. (2012). When public words are not data. In Heider, D. & Massanari, A. L. (Eds.), *Digital ethics: Research & practice*, p. 21. Peter Lang.

Dishman, L. (2019, January 25). Google Algorithms and Human Psychology: How Jigsaw Rescues Teens from ISIS Recruiters. *Fast Company*. https://www.fastcompany.com/90294876/how-jigsaw-is-using-ai-human-connections-and-adwords-to-fight-isis

Dixon, H. B. (n.d.). Telephone Technology versus the Fourth Amendment. https://www.americanbar.org/groups/judicial/publications/judges_journal/2016/spring/telephone_technology_versus_the_fourth_amendment/

Elangovan, N. (n.d.). Explainer: How the Criminal Law Reform Bill aims to fight crimes of the Internet age. *Today.* https://www.todayonline.com/singapore/explainer-how-criminal-law-reform-bill-aims-fight-crimes-internet-age

Forensic Experts Group. (2019, April 22). Staying Non-partisan—The Duty of an Expert. https://lawgazette.com.sg/practice/practice-support/staying-non-partisan-the-duty-of-an-expert/

Greenberg, A. (2017, June 3). Google's Clever Plan to Stop Aspiring ISIS Recruits. *Wired*. http://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/

Greenberg, A. (2018). Inside Olympic Destroyer, the Most Deceptive Hack in History. *Wired*. http://www.wired.com/story/untold-story-2018-olympics-destroyer-cyberattack/

Gstrein, O. J., Bunnik, A., & Zwitter, A. (2019, August 30). Ethical, Legal and Social Challenges of Predictive Policing. *Católica Law Review, Direito Penal*, *3*(3), pp. 77–98. Available at SSRN: https://ssrn.com/abstract=3447158

International Journal of Psychology. (2016). July 2016 Supplement, *51*(1), pp. 597–608.

Kennedy, M. (2012, December 14). Gary McKinnon will face no charges in UK. *The Guardian.* https://www.theguardian.com/world/2012/dec/14/gary-mckinnon-no-uk-charges

Keyworth, M., & Wall, M. (2016, January 8). The 'Bogus Boss' Email Scam Costing Firms Millions. *BBC News*. http://www.bbc.com/news/business-35250678.

Kipane, A. (2019, November 25). Meaning of profiling of cybercriminals in the security context. https://doi.org/10.1051/shsconf/20196801009

Kirwan, G. H. (2018). Dispelling the Pseudopsychology of Cybercrime. *Cyberpsychology, Behavior, and Social Networking*, *21*(2), pp. 71–72. doi: 10.1089/cyber.2017.29100.ghk

Kirwan, G. H., Fullwood, C., & Rooney, B. (2018). Risk Factors for Social Networking Site Scam Victimization Among Malaysian Students. *Cyberpsychology, Behavior, and Social Networking*, *21*(2), pp. 123–128. https://doi.org/10.1089/cyber.2016.0714

Kirwan, G., & Power, A. (2014). *Psychology of Cyber Crime: Concepts and Principles*. IGI Global.

Kushner, D. (2011, June 27). The Autistic Hacker. *IEEE Spectrum.* https://spectrum.ieee.org/telecom/internet/the-autistic-hacker

Liptak, A. (2018, June 22). In Ruling on Cellphone Location Data, Supreme Court Makes Statement on Digital Privacy. *The New York Times*. https://www.nytimes.com/2018/06/22/us/politics/supreme-court-warrants-cell-phone-privacy.html

Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration*, *42*(12), pp. 1,031–1,039. https://doi.org/10.1080/01900692.2019.1575664

Meyer, R. (2015, April 15). U.S. Supreme Court: GPS Trackers Are a Form of Search and Seizure. *The Atlantic*. https://www.theatlantic.com/technology/archive/2015/03/supreme-court-if-youre-being-gps-tracked-youre-being-searched/389114/

Neo, R. W. (n.d.). A look at key changes to Protection from Harassment Act. *Today.* https://www.todayonline.com/singapore/look-key-amendments-protection-harassment-bill

Power, A., & Kirwan, G. (2014). Cyberpsychology and New Media: A Thematic Reader. *Psychology Press*, pp. 3–8.

Steel, C. (2014). Idiographic Digital Profiling: Behavioral Analysis Based on Digital Forensics. *Journal of Digital Forensics, Security and Law*, https://doi.org/10.15394/jdfsl.2014.1160

Vohryzek-Bolden, M. (1997). Ethical Dilemmas Confronting Criminological Researchers, *Journal of Crime and Justice*, *20*(2), pp. 121–138, https://doi.org/10.1080/0735648X.1997.9721585

Whitty, M. T. (2018, February). Do You Love Me? Psychological Characteristics of Romance Scam Victims. *Cyberpsychology, Behavior, and Social Networking*, *21*(2), https://doi.org/10.1089/cyber.2016.0729

Wiederhold, B. (2014). The Role of Psychology in Enhancing Cybersecurity. *Cyberpsychology, Behavior, and Social Networking*, *17*(2).

Winerman, L. (2004). Criminal Profiling: The Reality Behind the Myth. https://www.apa.org/monitor/julaug04/criminal

Woman Scammed of S$1.4m Arrested; Police Believe She Stole Money from Her Company. (2019, October 12). *Today.* www.todayonline.com/singapore/woman-scammed-s14m-arrested-police-believe-she-stole-money-her-company?cid=h3_referral_inarticlelinks_03092019_todayonline

Who is an Expert Witness and How to Use Expert Evidence in Singapore. (2019, April 24). https://singaporelegaladvice.com/law-articles/expert-witness-use-expert-evidence-singapore/

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0012

# Chapter 12

# Optimise Defender's Advantage: Practical Approaches for Cybersecurity Defence

Chan Meng Fai* and Benjamin Goh†

*Government Technology Agency, Singapore*

**chan_meng_fai@tech.gov.sg*

†*benjamin_goh@tech.gov.sg*

## 12.1 Introduction

Director of the Federal Bureau of Investigation (FBI) Robert Mueller famously once said, "There are only two types of companies: those that have been hacked and those that will be." It was the year 2012 when Director Mueller said this, at the annual RSA conference. At the time, the cybersecurity threat was largely contained in the hands (or minds) of technical geeks and technology aficionados, and that cybersecurity, though an issue to be taken seriously (e.g., Morris Worm), was not an existential threat to be worried about.

Eight years later today, in 2020, the cybersecurity field has seen an explosion of interest and significantly impacts all sectors of the economy. Globally, we have seen cybersecurity attacks such as Heartbleed, WannaCry, and Petya capture the popular imagination on the perils of cybersecurity. Across governments, incidents such as the OPM hack—where the database of the US Government's Office of Personnel Management has been compromised by state-level actors—redefine

government espionage as we knew it. The now popular "Operation Olympic Games," as alleged in the exposé documentary "Zero Days," reinvented the options of diplomacy in the digital age. Seeing the breath of attacks that hit us daily, we await with bated breath—when is the next Mirai Botnet going to hit us, and what will happen if the internet is down for extended hours in our new digital economy?

All these events read like chapters in a movie script, but it is reality. Across all organisations, these are issues dealt with on a day-to-day basis. In Singapore, with the move towards adoption of digital technologies to become a "smart" nation, we create—not of our active willing—a valuable target in ourselves. However, fear should not get in the way of progress. There are concrete methods in which practitioners and policymakers can take to mitigate the seemingly impossible risk that is cybersecurity, and this chapter aims to provide concrete suggestions in pointing the way forward for cybersecurity defence.

## 12.2 Digitalisation of Companies, Cities and Government

Understanding the scope of cybersecurity requires situating the underlying trend of digitalisation that has driven companies, cities, and governments alike to adopt digital technologies to presumably deliver better quality services to its constituents. From Cloud technologies to process automation and artificial intelligence, we are living in an age where technology promises to do what was previously thought impossible—generate customers insights from purchase records in the matter of seconds, offer hyper-personalised browsing at the shopping mall, or deploy a city-wide smart lamppost system that can monitor a city's security activity from the comforts of a control room.

To pursue differentiation from each other, companies and governments are rapidly jumping onto the digitalisation bandwagon, sometimes even when the technology components are not fully understood properly. As a result, organisations are unlikely to possess good cybersecurity practices to protect the new equipment they have bought. To make matters worse, by adopting digital technologies aggressively, organisations may expose themselves to a larger "attack surface" (with more digital services, there are more points of entry for a cyber attacker to get into an organisation's systems). As such, mitigating the effect of cybersecurity attacks is always a balancing act, which requires all parties—from security engineers

to frontline delivery units—to make complicated trade-offs between reducing risk and keeping pace with business demands [Bailey *et al*., 2014].

Like other City and National Governments, Singapore is in the midst of a large-scale digital transformation. Dubbed the "Smart Nation," the Singapore Government has also embarked on the journey to integrate technology to improve lives, to prevent Singapore from falling behind other global cities [Ng, 2019]. This unstoppable grand scale of digitalisation effort is leading us to a new frontier where more technology will be adopted and weaved into everyday life. Among these technologies, two of them are exceptionally transformative: Cloud and Internet of Things (IoT).

### 12.2.1 *Cloud*

Cloud computing is the "delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet (the cloud) to offer faster innovation, flexible resources, and economies of scale" [Microsoft, n.d.]. Unlike the traditional approach in purchasing, integrating, and adopting "technology" within an enterprise's stack, the Cloud has provided organisations a way to use technology without owning its hardware; they only have to buy services on need-to basis. Depending on an organisation's assessment of its core functions, Cloud computing exists in models such as "Infrastructure-as-a-service" (where computing infrastructure is provided), "Platform-as-a-service" (where an entire platform of services is provided), and "Application-as-a-service" (where the entire application, including its component parts, are provided). This "everything-as-a-service" approach has enticed organisations to embark on their cloud journey with its flexibility and cost efficiency.

Not surprising, as part of its digitalisation efforts, the Singapore Government has also officially adopted a Commercial Cloud-First policy in June 2018. As part of the Commercial Cloud-First policy, a majority of Government Information and Communications Technology (ICT) systems will be migrated to the commercial Cloud over the next few years, allowing the Government to gain benefits such as access to best-in-class services for our engineers, lower hosting costs, and reduced system downtime [Ng, 2019].

As much as the benefits are available to harness, when companies start to migrate to the cloud, they will be faced with challenges to maintain

their security posture, since the threat landscape changes along with broader usage of cloud technologies [Diogenes & Ozkaya, 2019].

### 12.2.2 *Internet of things*

A "smart" city goes beyond a virtual realm of services and has to touch the physical lives and experience of its citizens. Within Singapore, this is done by connecting devices to the internet, through the Internet of Things (IoT). Put basically, the IoT connects any devices to the Internet, which includes everything from washing machines, watches, to wireless coffee makers.

Within the public and private sectors, there has been an increasing push for the use of sensors, IoT, and robotics to collect and analyse large amounts of sensor data to better help in operational planning and bring about greater manpower productivity.

The deployment of "smart" devices such as sensors in the physical world will continue to increase the risks and likely require novel threats mitigation efforts to counter them. In cyber-physical systems, the failure modes can be more widespread and coupled with common tasks and jobs, it leads to not only great opportunity but also greater threat [World Economic Forum, 2019].

## 12.3 Cyber-Attacks

Singapore's journey toward digitalisation, while impressive, has been similarly met with challenges, especially when serious attempts by malicious actors to compromise our digital systems have succeeded and made significant impact on the public sector. Many of these attacks have been documented publicly, and have ranged across government agencies in the defence, education, and health sectors.

These attacks are recent, and have also informed our response to deal with such risks. In February 2017, a cyber-attack occurred, resulting in 850 military personnel data being compromised [Ministry of Defence, 2017]. A few months later, in April, National University of Singapore (NUS) and Nanyang Technological University (NTU) were said to be attacked by state-sponsored attackers who attempted to steal classified research information [Tham, 2017]. One year later, in August 2018, one of the most significant attacks was reported. SingHealth, a national provider of healthcare services, experienced a data breach that resulted in 1.5 million patients' personal data

being stolen, while 160,000 of those had their dispensed medicines' records taken too [Tham, 2018]. Notable political office holders were among those whose data was leaked, leading to a significant revamp of public sector practices in data management. Just last year, in December 2019, suppliers to the Ministry of Defence logistics units had encountered cyber-attack through malicious malware, resulting in the personal information of 2,400 personnel of the Ministry of Defence (MINDEF) and Singapore Armed Forces (SAF) being leaked [Chong, 2019].

These examples and more show that Singapore is constantly under attack, and that cyber defence is a long-term endeavour of adapting and constantly evolving with the adversaries. Apart from traditional, off-the-shelf tools, attacks use "living off the land" adversarial activities (such as, use of PowerShell)[a] saw a significant increase in recent years. PowerShell-related cyber-attacks saw a massive increase of 1,000% [Symantec, 2019], and its resultant "fileless" attacks often bypass existing cybersecurity solutions.

However, solutions to cybersecurity go beyond "buying better technology solutions." After all, cybersecurity solutions are not only exorbitantly expensive but they also become obsolete rapidly. We will lay down four institutional reasons that defenders face when they try to protect systems, and thereafter suggest three guiding approaches on how institutions can approach build effective cybersecurity defence.

There is a now-cliched but ever-relevant trilemma in cybersecurity: security, functionality, and cost. Most organisations have to choose two out of the three priorities. As security practitioners, we sometimes worry too much about the worst-case scenarios—because they happen—but we also see four types of pressures that make it feel impossible to secure the enterprise environment.

### 12.3.1 *Desire to adopt latest technologies*

Organisations are breathlessly chasing after new technologies for digitalisation in order to keep themselves competitive. This could be done without a good understanding of their current environment as well as the

[a] PowerShell is a task-based command-line shell and scripting language that helps system administrators and power-users rapidly automate tasks that manage operating systems (Linux, macOS, and Windows) and processes.

one they intend to change to. As a result, security risks are not properly identified, resulting in huge security liabilities they took on from the start of their journey, without knowing it.

For example, in cloud adoption, more often than not, misconfiguration becomes the key ingredient to a potent cocktail of malicious events where attackers could easily exploit, due to the organisations' lack of understanding in the technology. This was perhaps most clearly seen in the recent Capital One breach, in which a misconfiguration in Cloud technology resulted in the loss of personal information for 100 million *Capital One* customers [McLean, 2019].

Shining "smart" IoT devices are also in many organisations' shopping list to help boost the enhanced technological context of their cyber-physical realm. However, more often than not, IoT devices are likely designed and built for their dedicated function or added on with internet connectivity for "smartness," without security in mind.

When organisations try to interface both their current technologies with newly acquired ones, the architecture approach to secure this new dynamic hits down hard on the IT architects and cybersecurity specialists, as the industry and academia continue to develop clarity on positioning and architecturing the concept of a secure cyber environment.

### 12.3.2 *Lack of clarity in usage of cybersecurity technology*

Cyber threat landscape continues to evolve and change rapidly. We are at a phase in which the risks of cybersecurity are well known, but solutions are unclear. This is where vendors play a big role in seeding Fear, Uncertainty and Doubt (FUD), by hammering inflated risks of potential incidents and breaches, and supplanting with promises of solving them with their own silver bullets. A classic example we have seen, for example, is how vendors market the scary scenario of attackers using Artificial intelligence (AI) and Machine Learning (ML), which breaks more "traditional" organisations into cold sweat.

Every season, in almost coordinated fashion, new buzzwords hit the town. There is always a "new generation" of something, and it promises to be the future of cybersecurity. Worried, senior leadership get confused and the "fear-of-missing-out" (FOMO) mentally will kick in—companies then engage in shopping sprees for silver bullets. However, while silver bullets are useful against the fiercest werewolves, they are put to use probably only once in a blue moon. Without clarity in cybersecurity objectives and

a real understanding of underlying systems, organisations are likely to over-spend on security, without protecting the areas that matter most.

Furthermore, as alluded to earlier, beyond the battle to defend enterprise networks, the explosion in cheap IoT devices that are incapable of built-in cybersecurity features continue to create an uphill task in defence. In some ways, the choice to adopt IoT at a large-scale is an implicit decision to minimise "cost" and maximising "usability" in the cost-usability-security trilemma. Organisations should not walk into a security nightmare. The Mirai Botnet has shown us that there are hundreds of thousands of unsecure (mostly IoT) devices that are unpatched or with default password, can be manipulated easily by any malicious attacker to mount an attack on any targeted entity. The 2016 Distributed Denial of Service (DDoS) attack on Dyn's East Coast servers brought massive disruption to Internet service in the US for a lengthy four hours, which is not something that any organisation wants to expose itself to, without careful thought of its consequences to the business.

### 12.3.3 *Limited cybersecurity manpower*

Not helping the situation, a projected 3.5 million cybersecurity positions around the world will be left vacant due to the lack of competent cybersecurity professionals [World Economic Forum, 2017]. As schools and training academies geared up to produce more cybersecurity professionals, they never seem to be fast enough to fill up the gaps. The complete process of attracting, developing, and retaining cybersecurity professionals also seems to be a draining task, which becomes a constant challenge resting on the organisation's shoulder.

In view of such a situation, one of the logical approaches would be to invest aggressively in cybersecurity technology to mitigate this workforce gap issue. However, if this is done at the extreme, the outcome might not be viable. Many organisations ended up purchasing the newest cybersecurity technology available, at times without the means to deploy them. Even when deployed, some organisations appear not to have the means to operate them functionally as part of their defence mechanism. Thus, these huge cybersecurity investments could then easily have ended up not achieving the intended outcome, and at worst giving organisations a false sense of security.

Without a competent workforce, even with the best cybersecurity technology adopted within the organisation, an organisation is unlikely to attain an effective cyber defence.

### 12.3.4 *Open-source and publicly available offensive tools and techniques*

While there are a number of known paid offensive tools out there, in recent years, cybersecurity professionals and enthusiasts are also developing and open-sourcing their own security testing tools, making them available for other like-minded professionals. This has resulted in the exponential increase of cybersecurity tools, which could be offensive in nature, and available publicly on platforms such as Github. These open-source tools can range from proof-of-concept (POC) exploit to a complete offensive testing toolkit. When one desires, they could be easily further weaponised to become a set of damaging arsenals that can deal serious impact to individuals or organisations.

Apart from that, collections of malicious software such as viruses, trojans, ransomwares are also available online. Learning offensive techniques such as obfuscation and evasion techniques are a lot easier compared to a few years back. Tutorials and documentation on attacks techniques are now obtainable through blogs, articles, videos as well as digital communities via platforms such as Facebook, Medium, Telegram, Slack.

With the wealth of offensive resources widely available, one should not be surprised that some of those high impact cyber-attacks executed involved these publicly available tools. This domain of growth is unlikely to slow down. The challenge remains on how much organisations could leverage on these tools to strengthen their defence.

## 12.4 Approaches to Optimise Defender's Advantage

In considering the zealous digitisation agenda, FUD in cybersecurity, along with managing the trilemma of functionality, security and cost, there are three approaches we suggest to maximise the defender's ability to neutralise or minimise the threat.

### 12.4.1 *Built-in end to end security defences—architecture and operating for digital age*

For any adoption of technology, how an organisation designs its architecture and deploys its systems has a direct impact on their ability to

deal with cyber threats—these are existential decisions, not an afterthought relegated to "operations." Organisations should take effort to understand the native security tools and control functions that come with these technologies, and use the opportunity to leverage them to construct a more secure environment. Most of the time, these native controls being set up properly will build up the intended baseline security defence against common attacks—if Capital One configured their Cloud environment properly, they could have been able to prevent the hacks that have significantly hurt their business reputation.

One approach to deploy systems securely is to go through a process of "hardening" them. "Hardening" refers to "the process of limiting potential weaknesses that make systems vulnerable to cyber attack" [Center for Internet Security, 2020, para. 2], and it forms the foundational cornerstone in establishing security baseline by (1) disabling all unnecessary technological functions; (2) limiting privilege administrative functions that can be used; (3) turn on built-in security functions. One approach cyber defenders frequently embark on would be to study and take reference from security baseline of technological systems and platforms derived by community-driven organisations like the Center for Internet Security (CIS).

After going through the hardening phase and when the technology is now in operation and "live," organisations should also review their operational procedures with regard to technology management *vis-à-vis* security. More often than not, the operational phase does not include sound processes to ensure the intended security controls remain relevant or enforced at the desired levels. As such, the effectiveness of the security controls will deteriorate over time. It is therefore very critical for organisations to include security checks and validation within their technology change management team, such as through a Change Advisory Board (CAB), to ensure the consistency and integrity of its established security standards.

### 12.4.2 *Adopt fit-for-purpose security technologies to secure the enterprise effectively*

The cybersecurity market will continue to grow, and along with that will be an increase in marketing intended technical buzzwords. For the next few years, it is almost inevitable for organisations to increase their

cybersecurity technology purchase activities in order to beef up their defence. Organisations will outsource cybersecurity work and solutions, but cybersecurity is rich and complex, and it is not uncommon to hear executives tell vendors to "come back with an end-to-end cybersecurity strategy for our next discussion." Organisations must always take ownership in determining their own cyber risk level. To stay safe, organisations have to remember never to outsource cybersecurity planning and strategy to others, especially at the negotiating table.

For cybersecurity planning, one of the approaches organisations can consider to adopt is to examine how their cyber defence measures fair against adversarial tactics. Organisations can structure their defence measures using a defence framework, such as NIST Cybersecurity Framework (NIST-CSF). The NIST-CSF provides a conceptual framework based on the following cyber defence functions: "Identify, Protect, Detect, Respond and Recover" [NIST, n.d.]. This can help organisations to visualise and prioritise their investment, and progressively roll out a proportional cyber defence avoiding potential panic buy that arises from FUD or FOMO.

To understand how attackers attack and achieve their objectives, organisations can take heed from adversarial frameworks such as MITRE ATT&CK, see Figure 12.1, or LockHeed Martin Cyber Kill Chain (Figure 12.2). They can then create an Attack-Defence Matrix (ADM) by mapping both attack-based and defence-based frameworks (Table 12.1), to structurally think through and have a more holistic understanding of


Figure 12.1. MITRE ATT&CK adversarial framework.

*Source*: Adapted from https://attack.mitre.org/resources/pre-introduction/.


Figure 12.2. Lockheed Martin Cyber Kill Chain adversarial framework.

*Source*: Adapted from https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html.

Table 12.1. An illustrative, non-exhaustive example of an Attack-Defence Matrix (ADM).

| | | NIST Cybersecurity Framework (NIST-CSF) | | | | |
|---|---|---|---|---|---|---|
| | | **Identify** | **Protect** | **Detect** | **Respond** | **Recover** |
| **Pre-ATT&CK** | *Pre-Exploitation*<br>Recon, Weaponise | Threat intelligence | Web Application Firewall (WAF) | | | |
| **ATT&CK** | *Exploitation*<br>Deliver, Exploit | Vulnerability Assessment (VA) tool | WAP, Firewall, Network Intrusion Prevention Systems (NIPS), Content Disarm and Reconstruction (CDR), Endpoints Protection Platform (EPP) | Security Information and Events Management (SIEM) | Endpoint Detection and Response (EDR) | Backup and recovery tool |
| | *Post-Exploitation*<br>Control, Execute, Maintain | | Firewall, NIPS | SIEM, Deception technology, Sinkhole | EDR, Security Orchestration, Automation and Response (SOAR) | Backup and recovery tool |

their cybersecurity deployment approach. This can allow organisations to identify relevant gaps and controls in place within their cyber defence technology solution [Sager, n.d.]. With a structured approach and proper planning, cybersecurity technology adoption will be more effectively examined and deployed.

### 12.4.3 *Leveraging open-source technologies for defence*

Organisations often obsess over proprietary and "secret" threat intelligence. However, like how hackers leverage open-source exploit there, there is a significant amount of open-source information for cybersecurity defence that is not often used within an organisation.

Knowing about the wealth of resources widely available online, cybersecurity defenders must leverage on this knowledge, such as Github or exploit databases, to enhance their awareness of cybersecurity threats. Understanding the offensive tools and techniques available will help organisations to stay relevant against potential cyber threats and to obtain early warning threat indicators so as to be in the position to tune up the necessary cyber defence mechanism to handle them.

Many organisations, both within the private and public sector, continue to underestimate the usefulness of open-source intelligence (OSINT), limiting the usage of these treasure troves of publicly available information. To begin to effectively tap on this knowledge, organisations should begin to engage the cybersecurity community around them. Engaging the ecosystem of cybersecurity practitioners, who thrive on open-source information (or OSINT) will help to overcome limited in-house resources and perspectives. Experience and insights from the community to utilise these offensive tools, as well as the defensive means could benefit all organisations to stack up their cyber defence efforts efficiently.

To go even further, instead of relying on the vendor to translate key insights to executives, it is useful for companies to begin building internal capabilities to effectively capture open-source technologies. For companies with sufficient scale and makes the business decision to invest in cybersecurity, they can go even further to build deep in-house capabilities, which have shown to generate high returns. Not only will they be able take control of core cybersecurity capabilities, they will also able to build products and customise solutions to its unique environment,

which is one of the best ways to tailor one's defensive posture for highest impact.

## 12.5 Conclusion

Getting cybersecurity right is a complicated task involving multiple stakeholders—for companies, its engaging not just the IT department, but also the frontline business functions, and every employee. For a huge bureaucracy like the Government, cybersecurity is at once about policies, system monitoring, threat intelligence, but also frequent testing, system architecting, and widespread cybersecurity awareness, which involves policy makers, public servants (e.g., teachers), and their respective vendors and partners. As Department of Homeland Security (DHS) Secretary Kirstjen Nielsen rightly remarked, "neither government nor industry is prepared to face the [cyber] threat alone" [Marks, 2019, para. 9].

In this chapter, we have addressed one part of the broader cybersecurity puzzle, and laid out some strategies for defenders to optimise their advantage as they navigate the forces that go against security. Security is, for better or for worse, a multi-stakeholder affair, and the forces against security often sound the loudest—everyone "cares" about cybersecurity, until they are forced to weigh the concerns of security with user complaints and ballooning security costs.

Nevertheless, it is useful to take heed at this juncture from Robert Mueller's prescient quote, "There are only two types of companies: those that have been hacked and those that will be." This is not a statement of concession and helplessness, but an objective assessment that in the digitalised world today, cybersecurity events *will* happen, in a constant negotiation (arms race, if you will), between the attackers and the defenders. As long as we stay vigilant, adaptive, and never be complacent, it is possible to digital and secure—and when we have mastered that sweet spot, we would have truly built a smart nation together.

## 12.6 References

Bailey, T., Miglio, A., & Wolf Richter, W. (2014). The rising strategic risks of cyberattacks. *McKinsey Quarterly.* https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-rising-strategic-risks-of-cyberattacks

Center for Internet Security. (2020). *CommunityForce uses CIS hardened images for its customers*. https://www.cisecurity.org/case-study/communityforce-uses-cis-hardened-images-for-its-customers/

Chong, C. (2019, December 21). 2 vendors for Mindef, SAF hit by malware; personal data of 2,400 staff could have been leaked. *The Straits Times.* https://www.straitstimes.com/singapore/healthcare-training-provider-hmi-institutes-server-infected-by-ransomware

Diogenes, Y., & Ozkaya, E. (2019). *Cybersecurity — Attack and Defense Strategies* (2nd ed). Packt Publishing.

Marks, J. (2019, March 19). The Cybersecurity 202: Government can't fight cyber threats alone, DHS secretary says. *The Washington Post*. https://www.washingtonpost.com/news/powerpost/paloma/the-cybersecurity-202/2019/03/19/the-cybersecurity-202-government-can-t-fight-cyber-threats-alone-dhs-secretary-says/5c9040651b326b0f7f38f1cc/

McLean, R. (2019). A hacker gained access to 100 million Capital One credit card applications and accounts. *CNN Business.* https://edition.cnn.com/2019/07/29/business/capital-one-data-breach/index.html

Microsoft. (n.d.). *What is cloud computing?* https://azure.microsoft.com/en-us/overview/what-is-cloud-computing/

Ministry of Defence. (2017). *Breach in MINDEF's I-net System.* https://www.mindef.gov.sg/web/portal/mindef/news-and-events/latest-releases/article-detail/2017/february/28feb17_nr

MITRE. (n.d.). *PRE-ATT&CK.* https://attack.mitre.org/resources/pre-introduction/

Ng, C. K. (2019). *Digital Government, Smart Nation: Pursuing Singapore's Tech Imperative*. https://www.csc.gov.sg/articles/digital-government-smart-nation-pursuing-singapore's-tech-imperative

NIST. (n.d.). *NIST Cybersecurity Framework.* https://www.nist.gov/cyberframework

Sager, T. (n.d.). *The Cyber OODA Loop: How Your Attacker Should Help You Design Your Defense*. https://csrc.nist.gov/CSRC/media/Presentations/The-Cyber-OODA-Loop-How-Your-Attacker-Should-Help/images-media/day3_security-automation_930-1020.pdf

Symantec. (2019). *2019 Internet Security Threat Report*. https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf

Tham, I. (2017, February 12). Hackers broke into NUS, NTU networks in search of government, research data. *The Straits Times*. https://www.straitstimes.com/singapore/hackers-broke-into-nus-ntu-networks-in-search-of-government-research-data

Tham, I. (2018, July 20). Personal info of 1.5m SingHealth patients, including PM Lee, stolen in Singapore's worst cyber attack. *The Straits Times.* https://www.straitstimes.com/singapore/personal-info-of-15m-singhealth-patients-including-pm-lee-stolen-in-singapores-most

World Economic Forum. (2017). *This is What the Future of Cybersecurity will Look Like.* https://www.weforum.org/agenda/2017/08/the-us-is-upping-its-game-against-cyber-attacks-but-the-security-industry-faces-a-huge-challenge

World Economic Forum. (2019). *Realizing the Internet of Things: A Framework for Collective Action*. http://www3.weforum.org/docs/WEF_Realizing_the_Internet_of_Things.pdf

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0013

# Chapter 13

# Hacking the Hacker's Psyche

Omer Ali Saifudeen

*National Security Coordination Secretariat, Singapore*

*Omer_Ali_SAIFUDEEN@nscs.gov.sg*

## 13.1 Introduction

The word 'hacker' has become inundated with so much myth from popular fiction that the true epistemology behind this word has been forgotten. Initially, a 'hack' was an innocuous term that technical students in the 1970s and 1980s from American colleges—such as the Massachusetts Institute of Technology (MIT)—adopted. Steeped in a playful culture of exploration, these students of technology sought to push the boundaries of a problem by taking it apart in the pursuit of innovative solutions that pushed the limits of existing knowledge [Thomas, 2003; Zhang, 2018]. A key ethos supporting this culture was that knowledge belonged to everyone; it was meant to be shared by all. Another belief of the original hackers was that they should be judged by their hacking skills when faced with a difficult problem, regardless of qualifications, age, race, or position in life [Levy, 1984]. However, around the 1980s when the personal computer became more available to the general public and not just prestigious universities and businesses, the 'hacker' label became a popular term. This label was ascribed to and

appropriated by an underground technical subculture that broke into computer systems, not to tinker and improve the system, but for personal gain, thrills, and at times, to get back at individuals they felt 'deserved it' [Power, 2016].

Today, cyber hacking has become normalised as an expected risk to be protected against. According to the Singapore Threat Report by US-based cyber-security company Carbon Black, in Singapore alone, at least 96% of organisations surveyed in the report suffered from at least one breach due to cyber-attacks during a 12-month period ranging from around October 2018 to October 2019. This was a marginal increase from the previous year's report. Geopolitical tensions and monetary gain were cited as the key reasons for the increased number of cyber-attacks in this study. Furthermore, respondents in the study said that the attacks were growing in complexity [Low, 2019]. According to the 2018 Singapore Cyber Landscape Report, cybercrime has been on the rise since 2016 [CSA, 2018].[a,b] Given the fact that more of our everyday activities, including work, have shifted online due to the COVID-19 epidemic, it seems that malicious hackers are capitalising on this lifestyle change and, in particular, the fear and confusion stemming from the epidemic. As such, there has been a rise in the number of cyber-attacks globally—particularly against hospitals, medical services, and those working from home [Boyden, 2020; Interpol, 2020].

This chapter will firstly, delve into the profiles of individuals who become hackers, deconstructing how they see themselves, their attitudinal characteristics, and their motivations. This sets the stage to examine if hackers are predominantly just criminals, pranksters, or if online psychopathy is a common trait among them. It will then take a deeper look into their recruitment process and in particular the developmental pathway of young hackers. This leads to an examination of the manifesto said to describe the hacker's ethos and the key attributes behind the hacker culture it represents. Finally, it will explore the more kinetic and deadly manifestations of this online threat stemming from collaborations and reciprocal influences between hackers not associated with any violent groups and those that are.

---

[a] 2016: 5,175 cases, 2017: 5,351 cases, 2018: 6,179 cases.

[b] Cybercrime in the report referred to online cheating, cyber extortion, and cases investigated under the Singapore Computer Misuse Act. 8.

## 13.2 Who are Hackers?

Hackers can be anyone, and take on any of the following personas discussed in this chapter. Some hackers could merely be playful pranksters whose main motivations are getting back at those in authority, the powerful, or someone deserving of 'punishment' for tangling with them. A favoured tactic would be to embarrass and publicly humiliate the target by revealing private information that is damaging to them. Getting a good handle on the thinking behind hackers comes from taking a closer look at whom they target.

For example, an Anonymous member called 'William' felt he was punishing individuals who deserved it, such as those who watch child pornography. William would lure men who watched child pornography by visiting the websites they frequent and signing up with usernames that would get their attention. Once he had their online contact, he would fake an official warning from the US Child Protective Services (CPS), stating that CPS had the IP address and computer ID of the men who accessed child pornography websites, etc. During the chats on MSN Messenger between him and the online paedophiles, William would post the fake warning on the MSN Messenger chat. William got a thrill from knowing that this terrified his victim. On 4chan,[c] harassing someone online in this way was colloquially termed by hackers as 'life ruin'. William relished in the thrill and sense of power he felt when he was able to do this [Olson, 2013, pp. 29–31]. Aside from the thrill of doing 'life ruin' on someone whom they felt deserved it, other reinforcing motivations for hackers include the sheer power they held over their victim once they submitted to their demands. Some of these hackers might not have held much influence or achievements in real life, but online, it was another matter [Olson, 2013, p. 42].

Some hackers, on the other hand, may also see themselves as pirates who plunder information for the sake of liberating it for the masses. This has been the cornerstone philosophy of many hackers, and they subscribe to the notion that information is meant to be free and shared by all. Despite differences in motivation, the one motivation that crucially unites all hackers is the thrill they experience during the process of hacking secure systems—identifying a loophole, manipulating tight security

[c] 4chan is an anonymous image board website that hosts forums on a variety of topics. Users are able to post anonymously.

systems, and finally, obtaining access to private data. The 'hunt' for buried and restricted information is what drives them, and when they get away with it with little real-world consequences, they experience an intoxicating feeling of invincibility and power having taken on the powerful. To learn the skills needed to do this, they would need to master social engineering to manipulate their victims. Such hackers acquire this ability easily by living in a subculture where 'taunting, lying and stealing' was the norm [Olson, 2013, pp. 25 & 30].

However, hackers are not a homogeneous entity; while the bulk of the hackers in Anonymous, for example, were single young men, their ranks have also included ordinary individuals. One example was Jennifer Emick, a married mother of four who was not part of any deviant subculture but resonated with some of the vigilante activism causes of Anonymous such as that against Scientology [Olson, 2013]. One of the things that could have attracted ordinary individuals such as Emick to join hacker groups like Anonymous was the hacker's mindset of not taking things for granted, their penchant for questioning authority or perceived rigid ways of doing things, and by doing so, challenging the norm [Olson, 2012]. In addition, the anonymity to be able to speak one's mind, as well as the seemingly egalitarian environment this anonymity was able to create was the other attraction. This is best exemplified by how a user of the early '2 Chan' image boards known as 'Shii' said; that by being unknown on these electronic messaging boards, one could counter the development of cliques, discourage those who joined for reasons of vanity, and seeking to become elites. Anonymity created a level playing field where the best argument won, and Shii illustrated this by citing how "on an Anonymous Forum … logic would overrule vanity" [Olson, 2013, p. 28]. This created the impression to potential members that hacktivist groups like Anonymous were not a collection of depraved misfits who were venting, but rather a group of rational thinking individuals who saw past coverups, fought injustice, and had the courage and underground skills to do so.

There is a need to distinguish between those who identify themselves with the hacker culture, and criminals who simply use the internet as a tool for crime. There are a few categories of cybercriminals. On one hand, most cybercriminals who participate in underground online communities tend to be lone actors with clean criminal records and no organised syndicate connections, maintaining a stable day job while undertaking illegal online activities occasionally on the side. Most were introduced to

the world of cybercrimes during their college days [Barysevich, 2017]. Although, there are also some hackers who belong to cyber-criminal syndicates or work for hacker networks that are sponsored by governments [Barysevich, 2017].

Hacktivists, on the other hand, want to correct a perceived wrongdoing and have a political agenda [Barysevich, 2017]. They feel they are on a moral crusade to get back at corrupt individuals, organisations, and countries whose power protect them from any repercussions [Osborne, 2006; Wallix, 2016]. For hacktivists such as Anonymous, this ability to hack into the realm of the powerful and correct any social injustices infused in them a strong belief in the power of technology. This also reinforced their existing notions of social injustice and ideas of right and wrong [John, 2017]. Some hackers label criminal hackers as "crackers" to differentiate them from hacktivists and legitimate ethical hackers hired by companies or governments out to test systems [Engel, 2019]. Having said this, it is also possible that hacktivists might try to justify their actions using altruistic ideals which are, in actuality, self-serving (e.g., for bragging rights). Other aspects of the cybercriminal include having a chaotic state of mind, a general intolerance to a diversity of opinion, being socially inept, and having insufficient social support for any inherent challenges [Saroha, 2014].

Common motivational drivers for hackers despite their differing profiles can be summarised into six main categories namely: peer recognition, power, curiosity, addiction, thrill-seeking to alleviate boredom, and political/personal vendetta acts. Hackers often expend more energy in their efforts than what they get in return [Abu Zuhri, 2016]. Despite this, hackers keep up their attempts because they are addicted to the sheer thrill of being able to penetrate an almost impregnable computer system and later, by bragging about their exploits online. Key factors that reinforce this motivation include curiosity over where potential loopholes in security might lay hidden and seeking it out, much like a treasure hunt. For such hackers, the appeal lies in the lure of opportunistic moments to steal information and being able to get away with it undetected. Other key factors include rationalising or justifying online theft in the name of broader values and ideals. The latter key factor was the case with Edward Snowden, who felt that he was essentially whistleblowing in the public interest by carefully selecting which documents to release [Gellman & Markon, 2013; Wallix, 2016]. Other hackers can simply be disgruntled individuals who want to get back at an organisation or specific personalities.

Outside of these motivations, for some hackers, hacking is nothing more than a tool for a criminal task as they are employed by criminal groups to steal data [Espinosa, 2017].

During the Black Hat USA conference in 2014 conducted by the cybersecurity firm Thycotic, a survey was done with 127 hackers. Results revealed that 51% of them said the "search for emotions" was their main motivation for hacking, and only a small percentage of them (18%) felt money was their main motivation. The survey report added that these findings reveal how "modern hackers are curious, they are bored and want to test out their abilities." Hackers also had a sense of invulnerability as 86% of those surveyed felt that they would not likely face consequences for their actions [Panda Security, 2015]. This lack of remorse, need for power and control over their victims, and impressions of grandiosity have also been observed in hacker profiles [Nuccitelli, 2014]. Hence, if there is some level of psychopathy in specific hackers, then these can fall into the broader category of Online Psychopaths who can be defined as those who are adept at using computer technologies to target, control, and manipulate people [Nuccitelli, 2014]. These can even include online sexual predators and scam artists [Nuccitelli, 2014].

### 13.2.1 *Online psychopathy*

Not all online psychopaths break the law or get caught by the legal system. On one hand, some might be high functioning sociopaths with successful careers and high social standing [Sartain, 2018]. On the other hand, their victims tend to be vulnerable individuals who are either ignorant, submissive, or psychologically distressed adults [Nuccitelli, 2014] who find a deep psychological need fulfilled by these online psychopaths.

According to Nuccitelli [2014], some of the affective and emotional characteristics that apply to all online psychopaths include the following:

1. Easily takes offence and is aggressive online when provoked
2. Pretends to show emotion and empathy in online forums
3. Easily bored
4. Does not admit online to feeling depressed
5. Does not experience fear when engaging in offline violence or online deviance

6. Finds communicating explicit details about lewd, violent, and graphic items online arousing
7. Sense of invincibility and is not deterred by online obstacles
8. Comfortable giving online criticism to others but rapidly becomes uncomfortable when on the receiving end of online criticism.

When these affective traits manifest into behaviour and online action, it usually entails a lot of online reputation management which is of great importance to online psychopaths [Nuccitelli, 2014].

Online psychopaths are skilled at utilising social engineering to trap their victims. Social engineering attacks by hackers stem from a keen knowledge of instinctive human behaviour. For example, clickbait attacks tap on the human tendency to "crane our neck to see an accident on the side of the road" [Zamora, 2018], or it could be observing browsing patterns to exploit a person's interests. The most common social engineering attacks occur on social media platforms, where they prey on an individual's concern over self-image and reputation. Hackers also exploit and manipulate their victim's trust. These hackers launch their attack by playing on this concern and while masquerading as someone the victim trusts, attempting to falsify concerning information. Another type of attack could be in the form of a phishing attack appearing to come from a legitimate source, or playing on sympathies for a particularly vulnerable profile one might have a soft spot for [Zamora, 2018].

## 13.3 How are Hackers Recruited?

Criminal groups have been known to hire hackers, and occasionally outsource some aspects of their hack [Merrill, 2015; Sussman, 2019]. A hacker wannabe first has to prove oneself. Consequently, highly skilled wannabe hackers may get noticed by other already proven hackers or criminal groups who may approach them for a job. Therefore, reputation management is key to getting recruited. Another way they are recruited is via the dark web. There are password-protected hacker communities on the dark web alongside websites where criminal groups recruit hackers for specific hacking jobs. Potential hackers may reach out to these recruiters, sometimes with recommendations from established hacker peers. This is almost akin to a normal job interview; there is the initial conversation with the recruiter except this is usually done over Skype without video, using

voice changing technologies and anonymised conversations using TOR[d] over hacker recruitment sites [Waddell, 2016].

Recruiters for hacking jobs might look for specific computer skills such as SQL injection, experience or skills with DDoS attacks, or competency in specific programming languages such as Perl, Python, or C. Recruiters also look for individuals with insider knowledge on the organisation they wish to hack [Waddell, 2016]. The big difference in such recruitment is there is a lot of caution involved as there is the ever-present danger of law-enforcement undercover operatives seeking to flush hackers out. Usually, precautions taken involve a probationary period and specific tasks the hacker on probation must achieve in order to prove oneself [Waddell, 2016].

However, it is important to note the potential danger hackers place themselves in—by getting recruited by terrorist groups or foreign governments through blackmail. This is because once they accept a hacking job for money, they are essentially 'owned'. In some instances, this information is used by governments or terror groups at times to coerce these hackers to work for them. In other instances, these governments or groups use ideology or play on the ego of the hacker to recruit them [Davis, 2018].

## 13.4 Developmental Pathway of Young Hackers

For very young hackers such as those in their tweens, the pathway into the world of hacking starts with a simple hack, and escalates once these youth learn they can make money out of it. They relish how being a hacker makes them appear 'cool' among their peers. Through interaction with older or more seasoned hackers and criminals online they become savvier and more immersed into hacking culture [Yusof, 2016].

Take the case of the hacker known as 'Alok'[e] who, in his early teens, was already a hacker making thousands of dollars on the Dark Web as part of a hacker collective. Alok's beginnings into hacking started out with the expected thrill-seeking adrenalin rush that came from breaking into a well-protected system, a validation of his abilities to do something

[d]TOR is a web browser that redirects web requests via a random string of servers to mask the origin of the requests.

[e]'Alok' is not his real name, and he is believed to be Indian.

seemingly impossible, as well as in proving his skill to other hackers. Alok's first hack was finding a vulnerability and breaking into the user data of a start-up in Bangalore. He reported this vulnerability to the company and was rewarded with a free t-shirt. Importantly, he got away with no repercussions for his actions. This subsequently emboldened him to continue hacking [John, 2017].

Alok spent around three years engaging in activities such as theft and trading of credit card information. Alok, however, was not the leader of a hacker clan but was rather a follower who earned money in bitcoins, taking care in hiding his wealth from his parents. However, as he matured, he came to a point where his youthful exuberance had given way to practical realities, leading to him question if he really wanted to continue as a hacker [John, 2017]. Hackers are known to get easily bored if the activity does not invigorate them over time. This suggests some might just grow out of this subculture if the community and causes they advocate no longer inspire, fulfil any of their thrill-seeking needs, or they simply grow out of it as other life concerns and motivations come to dominate.

According to the Singapore Cybersports & Online Gaming Association (SCOGA), in Singapore, there have been youths who have supported the values of hacktivist groups, conducted high-profile cybercrime acts, or have been involved in other forms of troubling behaviour online, such as misrepresenting their age to take part in online betting. Most of the at-risk young people referred to SCOGA for counselling come from vulnerable families. Common reasons for this vulnerability stems from family tensions and parental neglect, which makes them turn to online communities. It could also be due to struggles in school, or the inability to conform to social norms that lead them to adopt alternative and radical values online. SCOGA reaches out to many game enthusiasts, and every year counsels around five to 10 young people in the age group of 11 to 21 [Yusof, 2016].

## 13.5 Hacker's Manifesto and Culture

In 1986 in the US, Loyd Blankenship, known by his hacker nickname 'The Mentor', wrote an essay called *The Conscience of a Hacker*, or the *Hacker Manifesto*, when he was in prison after being arrested for hacking. It was later published in the underground hacker "ezine," *Phrack*, and became highly popular within the hacker subculture, and was eventually adopted as their creed. Blankenship's essay laid out the

ethical foundation for hacking, and emphasised how the whole idea behind hacking went beyond selfish desires, exploitation, or doing harm. Hacking, in his idealised version, was meant to expand the minds of people and keep the world free [*Phrack Magazine*, 1986]. A rhetorical analysis of the essay in the *Phrack* magazine lists out the key characteristics of the identity hackers claimed to espouse and the justifications for their actions.

(1) *The Hacker as the Victim*: Hackers claim to be a misunderstood elite group that is vilified for being smarter and more talented than the slower masses who label them negatively, and punishing them for not being like the rest or following social norms.

**Excerpt from *Hacker Manifesto*:**

"… I am a hacker, enter my world …
Mine is a world that begins with school … I'm smarter than most of the other kids, this crap they teach us bores me …
Damn underachiever. They're all alike.

I'm in junior high or high school. I've listened to teachers explain for the fifteenth time how to reduce a fraction. I understand it. 'No, Ms. Smith, I didn't show my work. I did it in my head …'
Damn kid. Probably copied it. They're all alike …."

[*Phrack Magazine*, 1986].

(2) *The Computer Liberates*: The computer, to the hacker, is an unbiased and near-perfect instrument that unlocks a world in which they finally feel they belong and are accepted for who they are.

**Excerpt from *Hacker Manifesto*:**

"I made a discovery today. I found a computer. Wait a second, this is cool. It does what I want it to. If it makes a mistake, it's because I screwed it up. Not because it doesn't like me …

Or feels threatened by me …
Or thinks I'm a smart ass …
Or doesn't like teaching and shouldn't be here …

Damn kid. All he does is play games. They're all alike.

> And then it happened … a door opened to a world … rushing through the phone line like heroin through an addict's veins, an electronic pulse is sent out, a refuge from the day-to-day incompetencies is sought … a board is found.
>
> 'This is it … this is where I belong …'
>
> I know everyone here … even if I've never met them, never talked to them, may never hear from them again … I know you all …"
>
> [*Phrack Magazine*, 1986]

(3) *The Hacker as the Freedom Fighter/Rebel*: The hacker claims to be the rebel against a corrupt, bigoted, hypocritical, and suppressive elite that rules over them and the masses. According to the hacker, this elite—despite being responsible for the injustices, unfairness, and exploitation of society—ironically view hackers as criminals.

**Excerpt from *Hacker Manifesto*:**

> "We explore … and you call us criminals. We seek after knowledge … and you call us criminals.
>
> We exist without skin color, without nationality, without religious bias … and you call us criminals.
>
> You build atomic bombs, you wage wars, you murder, cheat, and lie to us and try to make us believe it's for our own good, yet we're the criminals.
>
> Yes, I am a criminal. My crime is that of curiosity. My crime is that of judging people by what they say and think, not what they look like. My crime is that of outsmarting you, something that you will never forgive me for.
>
> I am a hacker, and this is my manifesto. You may stop this individual, but you can't stop us all … after all, we're all alike."
>
> [*Phrack Magazine*, 1986].

In 1984, Steven Levy wrote a book that explored the hacker culture in which he described the 'Hacker Ethic'. Key tenets [Levy, 2001] of this ethic began with the ideal that everyone should have access to computer systems that can reveal information about the world. It went onto other ideals such as everyone should have free access to all information; to never trust anything from authority; encourage decentralisation;

Table 13.1. Summary of Key Motivations espoused in *Hacker Manifesto.*

| | Motivation | Inferred Claims |
|---|---|---|
| 1. | The Hacker as the Victim | Hackers claim to be a misunderstood elite group vilified for being intellectually gifted and not following social norms. |
| 2. | The Computer Liberates | The computer is an unbiased instrument that connects them to a world they belong, are treated equally and empowers them. |
| 3. | The Hacker as the Freedom Fighter/Rebel | Hackers claim to be rebels against a corrupt, bigoted, hypocritical, and suppressive elite that is controlling them by controlling information. |

believe that art and beauty can be created through the computer and hence hacking is an art form; and finally, trust that everyone's life can be improved through computers.

Essentially, the hacker is personified as a modern-day 'Robin Hood' who rebels against authority figures who restrict access to information that the public should be aware of [Levy, 2001]. Over time, as personal home computers became commonplace and networked, this hacking culture and belief also spread to technical epicentres in Europe such as Germany, as well as the rest of the world. For example, the German white hat hacking group known as 'Chaos Computer Club' is one of the oldest in the world and has its beginnings in 1981. In many ways, this ethical hacking group is a synthesis between political activism aimed at fighting for government transparency, privacy, and the removal of restrictions in sharing information and technical hacking [Brooke, 2011] (Table 13.1).

### 13.5.1 *Hacker's hangouts*

The online spaces where hackers congregate need to be conducive to their culture as well. For example, while 4chan—the platform where Anonymous started—was an environment rife with porn, and all kinds of depravity and nasty vulgar in-group hacker lingo, it was also an online space that promoted egalitarianism and created a space where creative ideas could flourish. It made its users feel as if they were in a secret exclusive club where they belonged [Olson, 2013].

A similar online psychopathy, and shared hacker networks, culture, and identity could also explain why hackers with no affiliations to violent groups (secular hackers) have been found to collaborate at times with hackers for violent groups.

### 13.5.2 *Collaborations between secular hackers and hackers for violent groups*

Taking the case example of hackers for jihadi groups, these hackers at some point had their beginnings in mainstream hacker culture. The Algerian hacker group Moujahidin Team (El Moujahidin) that defaced the Air France website with pro-Algerian nationalistic rhetoric on 30 March 2015 had issued statements claiming grievance over the killing of Muslims, as well as airing historical grievances over France's colonial history in Algeria and treatment of Palestinians [Staff, 2015]. While citing religious justifications, this group openly displayed and made references to punk rock culture which in essence was at odds with strict Islamic religious ideals. For example, the group often used phrases like "Smoke Weed Everyday" and referenced rap music in their posts. The group, while claiming to follow the "Mujahideen mentality," said they did not follow any particular jihadi group. However, they have also appeared with other major jihadi slanted hacking groups (e.g., AnonGhost team and Algerian Cyber Army) while claiming to be one of them as well. At some points, the group has even claimed to perform cyber-attacks not because of any ideological motivation but 'just for fun', similar to how Lulzsec claimed it was doing such attacks for 'lulz' (hacker lingo that meant hacking for fun and laughs) [Staff, 2015].

Moujahidin Team's attacks had also gone beyond Algeria to include attacking the Indonesian Jambi City General Elections Committee website, domains in Belgium, and Fila sportswear's Russian domain. Hence, the targets of their attacks were starkly different from that of typical jihadi hacker groups such as the attack conducted by the ISIS "CyberCaliphate," when it hacked the United States Central Command (CENTCOM)'s Twitter account [Staff, 2015].

An example of a secular Hacker who became an ISIS member was the late Junaid Hussain, leader of the ISIS CyberCaliphate who went by the jihadi name Abu Hussain al-Britani. He used to be the leader of the secular hacking group known as 'TeaMp0isoN' and went by the hacker

nickname 'TriCk'. When Hussain later joined ISIS, he was also known to be recruiting hackers for the ISIS CyberCaliphate that he led. At one point, Hussain sent support for the Ferguson rioters[f] in the US, and promised to send militants down to help them if they pledged allegiance to ISIS [Hall, 2014]:

> "We hear you and we will help you if you accept Islam and reject corrupt man-made laws like democracy and pledge your allegiance to Caliph Abu Bakr and then we will shed our blood for you and send our soldiers that don't sleep, whose drink is blood, and their play is carnage."
>
> [Hall, 2014].

Hussain, when he was with 'TeaMp0isoN', fought against Islamophobia, Israel's handling of Palestine, and the issues in Kashmir, but was by no means violent. He was much loved by his members in 'TeaMp0isoN' and his hacker partners in Anonymous. Hence, it came as a shock to them when he eventually joined ISIS. This change towards militancy and violence came about after he was imprisoned for his hacking offences. Apparently, ISIS did not have to aggressively look for people such as Hussain to recruit. Instead, he sought them out. ISIS knew how to exploit him by giving him leadership, prestige, and money. Hussain, in turn, hacked effectively for the ISIS CyberCaliphate and led this group. An Anonymous member once reported how the Syrian Electronic Army and ISIS CyberCaliphate often sought to learn from secular hacker groups like Anonymous on target handling and how to manipulate the audience watching their actions. After Hussain's death at the hands of the US military, the capabilities of the ISIS CyberCaliphate degraded significantly [Murphy, 2015].

An important point to note is that hackers like Hussain did not start out hacking for Islam. Even as they eventually hacked for issues important to the Muslim world, they strongly identified with the overall secular hacker identity. Take the example of Muslim hacker '0xOmar' who conducted cyber-attacks on Israel in an Anonymous group operation called #Op-Israel. He knew Junaid Hussain at one point and was friends with him. However, '0xOmar' eventually ended up hacking ISIS websites

[f]The massive riots in Ferguson, Missouri, US, began after the fatal shooting of African American Michael Brown by a white police officer on 9 August 2014.

for Anonymous, and claimed in an interview that he "fights for peace in this world" [Paganini, 2016; Haas, 2012].

In 2015, when Anonymous started their campaign called #OpIsis against ISIS by hacking their sites, there were Christian, Muslim, and Jewish hackers from Anonymous working together against a common enemy [Ou, 2016]. A voice in the Anonymous video said:

> "We are Muslims, Christians, Jews, we are hackers, crackers, Hacktivist, phishers, agents, spies, or just the guy from next door,"
>
> [Ou, 2016].

Cooperation between jihadi hacking groups and other hackers was also seen when the Cyber Caliphate connected and co-hacked with the hacking group known as Lizard Squad[g] [Listes, 2018]. A few of Lizard Squad's members apparently also identified themselves with Cyber Caliphate on their Twitter accounts [Recorded Future, 2015]. In January 2015, Lizard Squad also hacked the Malaysian Airlines website, and posted the image of the Lizard Squad mascot and wrote "Hacked by Lizard Squad—The Official Cyber Caliphate" on the page [Newton, 2015].

What we could be seeing is how jihadi hacker collectives are building capacity through influencing and cooperating with other secular hacker collectives. While such collaborations and mutual influences between secular hackers and hackers for violent groups might seem to be at the fringes, the conduits, networks, and opportunities for such collaborations exist because both ascribe to the same hacker psyche/culture and share common hacker associations/networks from the past. We should be mindful of this reciprocal contagion effect.

## 13.6 Conclusion

Often than not, hackers are not the monolithic entity they are perceived to be. At one end of the spectrum that speaks of their origins, you have playful tinkers who sought to find unorthodox solutions to problems. This later morphed into anti-establishment hacktivists who claimed to fight for

[g]Known members of Lizard Squad include Vinnie Omari, Julius Kivimaki, Zachary Buchta, and Bradley Jan Willem van Rooy [Listes, 2018].

information freedom and sociopolitical causes. At the darker end of the spectrum, you have online psychopaths who satisfy their perverted needs for control and power. Finally, you have criminals who exploit the internet for gains, and hackers who see themselves as another arm of violent groups. What links all of them could be a hacker counter-culture that view computers as a liberating medium that empowers them, an empowerment and sense of control they could not have otherwise achieved in the real world. Hence, any efforts to bring someone out of the hacker culture first needs to examine which part of the spectrum they fall into. Following this, offering alternatives that either redirect their abilities towards doing good (e.g., via ethical hacking) or creating healthy, real-world options that address their specific motivations to give them a sense of agency and empowerment might be the best way to 'hack' into their psyche and send them down a better path.

## 13.7 Acknowledgement

The views expressed in this chapter are the author's only and do not represent the official position of the National Security Coordination Secretariat (NSCS).

## 13.8 References

Abu Zuhri, F. (2016, June 18). *The Profile of a Cybercriminal*. Digital Forensics Magazine. https://digitalforensicsmagazine.com/blogs/wpcontent/uploads/2017/05/The-Profile-of-Cybercriminal.pdf

Barysevich, A. (2017, November 28). *Inside the Mind of Cybercriminals*. Recorded Future. https://www.recordedfuture.com/cyber-criminal-profiling/

Boyden, P. (2020, April 9). *Remote Worker Cyber-attacks Increase Amid the COVID-19 Pandemic*. Fraud Watch International. https://fraudwatchinternational.com/all/remote-worker-cyber-attacks-increase-amid-the-covid-19-pandemic

Brooke, H. (2011, August 24). *Inside the Secret World of Hackers*. The Guardian. https://www.theguardian.com/technology/2011/aug/24/inside-secret-world-of-hackers

Cyber Security Agency (CSA). (2019, June 1). *CSA: Singapore Cyber Landscape 2018*. https://www.csa.gov.sg/news/publications/singapore-cyber-landscape-2018

Davis, D. (2018, June 29). *How Cyber Terrorists Recruit Hackers—and How They Can Resist*. PR Newswire. https://www.prnewswire.com/news-releases/how-cyber-terrorists-recruit-hackers--and-how-they-can-resist-300285718.html

Engel, K. L. (2019, October 14). *Social Justice Hackers: Tech Patriots or Pretenders?* https://www.whoishostingthis.com/blog/2019/09/12/hacking-activists/

Espinosa, N. (2017, February 16). *Hacker Mentality: Who Are They and Why Are They Targeting Me?* Smartfile. https://www.smartfile.com/blog/hacker-mentality

Gellman, B., & Markon, J. (2013, June 10). *Edward Snowden Says Motive Behind Leaks was to Expose 'Surveillance State*'. Washington Post. https://www.washingtonpost.com/politics/edward-snowden-says-motive-behind-leaks-was-to-expose-surveillance-state/2013/06/09/aa3f0804-d13b-11e2-a73e-826d299ff459_story.html

Haas, S. (2012, January 18). *'0xOmar' demands Israeli apology*. Ynetnews. https://www.ynetnews.com/articles/0,7340,L-4176436,00.html

Hall, J. (2014, November 26). *ISIS Supports Ferguson Protesters and Pledge to Send Over Soldiers.* Daily Mail. https://www.dailymail.co.uk/news/article-2850442/We-hear-help-ISIS-tweets-support-Ferguson-protesters-reject-corrupt-man-laws-like-democracy.html

Interpol. (2020, April 4). *Cybercriminals Targeting Critical Healthcare Institutions with Ransomware*. https://www.interpol.int/en/News-and-Events/News/2020/Cybercriminals-targeting-critical-healthcare-institutions-with-ransomware

John, N. (2017, October 16). *Hacking the Mind of a Hacker: ET Decodes the Psychology of a Cybercriminal*. The Economic Times. https://economictimes.indiatimes.com/tech/internet/hacking-the-mind-of-a-hacker-et-decodes-the-psychology-of-a-cybercriminal/articleshow/61101532.cms

Levy, S. (1984). *Hackers: Heroes of the Computer Revolution.* Anchor Press: Doubleday.

Levy, S. (2001). *Hackers: Heroes of the Computer Revolution.* Harmondsworth: Penguin Books.

Listes, L. (2018, August 1). *The Top 10 Most Famous Hacker Groups: Les Listes*. Les Listes. https://leslistes.net/top-10-famous-hacker-groups/

Low, Y. (2019, October 1). *High Volume of Cyber Attacks in S'pore in Past Year, Mostly Ransomware: Report.* TodayOnline.https://www.todayonline.com/singapore/high-volume-cyber-attacks-singapore-year-ransomware-chief-attack-mode-report

Merrill. J. (2015, June 7). *Major Effort Needed to Tackle Criminal Gangs Using 'Hackers for Hire'*. https://my.independent.co.uk/news/uk/home-news/major-international-effort-needed-to-fight-criminal-gangs-using-hackers-for-hire-says-anti-10302198.html

Murphy, L. (2015, December 19). *The Curious Case of the Jihadist Who Started Out a Hacktivist.* Vanity Fair. https://www.vanityfair.com/news/2015/12/isis-hacker-junaid-hussain

Newton, J. (2015, January 27). *Malaysia Airlines Website Hacked by Lizard Squad over MH370 Disappearance.* Mail Online. https://www.dailymail.co.uk/news/article-2926315/Malaysia-Airlines-website-targeted-hacker-group-Cyber-Caliphate.html

Nuccitelli, M. (2014). *Online Psychopaths.* https://docs.google.com/file/d/0B9oaR2swPnLbUk9xYVM3cG1LLWM/edit?pli=1.

Olson, P. (2012, July 31). *Exploding The Myth Of The 'Ethical Hacker'*. Forbes. https://www.forbes.com/sites/parmyolson/2012/07/31/exploding-the-myth-of-the-ethical-hacker/#5c905c7333ea

Olson, P. (2013). *We are Anonymous: Inside the Hacker World of LulzSec, Anonymous, and the Global Cyber Insurgency*. Back Bay.

Osborne, M. (2006). Chapter 12. In *How to Cheat at Managing Information Security.* Syngress.

Ou, J. (2016, January 19). *Hacking Group Anonymous Takes Down ISIS Websites, Social Media Accounts*. The Straits Times. https://www.straitstimes.com/world/middle-east/hacking-group-anonymous-takes-down-isis-websites-social-media-accounts/

Paganini, P. (2016, September 4). *Hacker Interviews—0xOmar (@0XOMAR1337).* Security Affairs. https://securityaffairs.co/wordpress/50938/hacking/hacker-interviews-0xomar.html

Panda Security. (2015, October 15). *Inside the Mind of a Cybercriminal: What is He Looking for and Why has He Chosen Your Business?* https://www.pandasecurity.com/mediacenter/security/inside-mind-cybercriminal/

Phrack Magazine. (1986, September 25). *The Mentor*. http://www.phrack.org/issues/7/3.html#article

Power, K. (2016, August 17). *The Evolution of Hacking*. Tripwire. https://www.tripwire.com/state-of-security/security-data-protection/cyber-security/the-evolution-of-hacking/

Recorded Future. (2015, April 2). *Cyber caliphate: ISIS plays offense on the web*. Recorded Future. https://www.recordedfuture.com/cyber-caliphate-analysis/

Saroha, R. (2014). In *Profiling a Cyber Criminal. International Journal of Information and Computation Technology 4*(3), pp. 253–258. https://www.ripublication.com/irph/ijict_spl/ijictv4n3spl_06.pdf

Sartain, A. (2018, June 14). *High-Functioning Sociopaths and How to Spot Them.* Mindcology. https://mindcology.com/mental-health/high-functioning-sociopaths-spot/

Staff, V. (2015, 3 April). *Air France Cyberattack: Who is the Moujahidin Team and Why are they Waging Cyber-jihad?.* IBTimes. https://www.ibtimes.co.uk/air-france-cyberattack-who-moujahidin-team-why-are-they-waging-cyber-jihad-1494807/

Sussman, B. (2019, May 5). *'Do YOU Want to Get Rich?' Hackers Are Hiring and Looking for Business.* SecureWorld. https://www.ripublication.com/irph/ijict_spl/ijictv4n3spl_06.pdf

Thomas, D. (2003). *Hacker Culture.* Univ. of Minnesota Press.

Waddell, K. (2016, March 2). *How Hackers Recruit New Talent.* The Atlantic. https://www.theatlantic.com/technology/archive/2016/03/how-hackers-recruit/471729/

Wallix. (2016, May 31). *The Psychology of the Cyber Criminal—Part I*. http://blog.wallix.com/the-psychology-of-the-cyber-criminal

Yusof, Z. M. (2016, October 3). *How Tweens Get Recruited to Become Hackers*. The New Paper. https://www.tnp.sg/news/singapore/how-tweens-get-recruited-become-hackers

Zamora, W. (2018, August 17). *Hacking Your Head: How Cybercriminals Use Social Engineering.* https://blog.malwarebytes.com/101/2016/01/hacking-your-head-how-cybercriminals-use-social-engineering

Zhang, N. (2018, June 28). *Hacker Culture—History and Influences | TEDxSMICSchool.* Youtube. https://m.youtube.com/watch?v=9cr1XZlEir0

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0014

# Chapter 14

# Humans as the Weakest Link in Maintaining Cybersecurity: Building Cyber Resilience in Humans

Pamela Goh

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*melamerc_pam@hotmail.com*

## 14.1 Introduction

Concerns in the area of cybersecurity have recently been placed under the spotlight, particularly so when 2017 saw a substantial number of cyberattacks and security breaches across the world [Graham, 2017; Leech, 2017]. In April 2019, the passwords and photos of millions of Facebook users around the globe were leaked by third-party companies [Osborne, 2019]. In Singapore in January 2019, the personal details of 14,200 HIV-infected individuals were also illegally obtained by an unauthorised person, to which this information could still eventually be disclosed publicly and unlawfully online today [Ministry of Health, 2019]. In fact, the incessant news of successful cyberattacks across the globe, such as in the aforementioned examples, are a stark reminder that cyber threats are real and here to stay. The disruption and costs associated with cyber threats have been shown to increase exponentially over time, rendering such threats as one of the major concerns for many developed countries [Drzik, 2018; World Economic Forum, 2018].

When successful, the consequences of cyberattacks can be profound and manifold. At the individual level, for instance, confidential and sensitive data can be compromised, financial losses can occur, and essential operations can be disrupted [Accenture & Ponemon Institute LLC, 2017; Tham, 2017a]. On the wider, macro level, those with malicious intents seeking access into computer and network systems can compromise and cause the collapse of "critical national infrastructures (such as energy, transportation, government operations) or to coerce or intimidate a government or civilian population" [Lewis, 2002, p. 1]. With the detrimental effects of cyberattacks, it is vital that these cyber threats are managed effectively. This chapter will serve to shed light on the weakest player in the maintenance of cybersecurity—humans.

## 14.2 Human Behaviours as the Weakest Link in Cybersecurity

Cyberattacks can be perpetuated via two means: (1) system-centric approach, where perpetrators exploit the technical vulnerabilities of a computer or network system to conduct an attack; and (2) user-centric approach, where negligence or mistakes of the computer users facilitate the execution of cyberattacks [Neupane, Rahman, Saxena, & Hirshfield, 2015]. However, successful cyberattacks in reality are often a result of the latter, in which human errors rather than technological shortcomings are the main cause of concern [Kelly, 2017; Tasman-Jones, 2016]. According to Symantec, 97% of malware attacks in 2016 targeted people and their poor online behaviours, with only the remaining 3% attributed to actual flaws in the network security system itself [Bennett, 2017].

As a result, perpetrators commonly employ social engineering; the method of hacking humans—rather than the system—by exploiting poor cyber behaviours to gain a backdoor in computer systems and networks [Choo *et al*., 2016]. It encompasses deceiving and psychologically manipulating the victims into divulging confidential information and/or getting them to perform certain actions that facilitate the execution of cyberattacks. For instance, the social engineering technique of "phishing" works because of the meticulously crafted email messages that encourage recipients to click on the weblinks or download attachments that are malicious in nature [Lord, 2017; Tham 2017]. Should one exercise

caution, successful phishing attempts may be avoided, as poor human behaviour in the cyberspace forms the core of why many such cyberattacks are successful in the first place [Bennett, 2017; Fallows, 2011; Hadlington, 2017]. As what Andy Waterhouse, EMEA Director at RSA Security [cited in Bennett, 2017, p. 12], had commented:

> "It is not just about silly errors but often a lack of training and understanding of the implications of clicking on a malicious link, going to a risky website or even setting up a service on a public cloud service without looking at the security implications."

The endeavour to combat one's susceptibility to cyberattacks involve firstly, a long process of understanding the specific attributions that contribute to this susceptibility, as well as the necessary follow-up actions needed to be done to manage the threats. Approaches to mitigate cyberattacks, therefore, go beyond the protective capabilities of sophisticated technological solutions [Conteh & Schmick, 2016; Goldman, 2013] such as antivirus software and firewalls, and have to include understanding human-centric measures.

## 14.3 Understanding Human-Centric Measures

### 14.3.1 *Importance of good cyber hygiene to manage cyber threats*

Perpetrators are constantly searching for the weakest link in the computer or network system, in order to gain quick and easy but unauthorised access into these areas [Ashiq, 2015]. Humans are unfortunately very much the weakest link in cybersecurity, because of their risky behaviours in the cyberspace [Vishwanath, 2016]. A 2017 survey conducted by the Cyber Security Agency of Singapore (CSA), for instance, revealed that many people exhibit poor behaviours (e.g., using the same password for both personal and work accounts, not using two-factor authentication, not installing security on mobile phones, not running virus checks for files and devices before opening them) on online platforms that put themselves and their organisations at risk of cyberattacks [CSA, 2017b].

If poor human cyber behaviours are indeed the main cause of successful cyberattacks, then managing these behaviours should reduce

one's vulnerability towards cyberattacks. To do so, there is a need to reduce risky online behaviours, in addition to improving one's cyber hygiene behaviours.

### 14.3.2 *The value of cyber hygiene behaviours*

Protective cyber behaviours are known as cyber hygiene behaviours. Just as how people brush their teeth and take showers to maintain their basic hygiene level and health, cyber hygiene practices are similar, except that they take place in the cyberspace [Symantec Corporation, 2017]. Cyber hygiene revolves around the implementation of cybersecurity 'best practices' to protect and maintain one's online 'safety' and 'health' whilst using the computer and internet. It focuses on individual responsibility to perform the identified best practices, rather than depending on technical protection measures against cyber threats [Aldorisio, 2018; Ashiq, 2015]. Some examples of cyber hygiene practices include using anti-virus software, activating firewalls, updating operating systems, wiping the files on a hard drive, and adopting the use of complex and unique passwords [Symantec Corporation, 2017].

The many years of effort to educate people on these protective measures[a] have seemingly produced limited results because people are generally still not undertaking adequate cyber hygiene practices. The engagement of sufficient cyber hygiene practices serves to enhance one's cyber resilience, in which their exposure—and therefore—likelihood of falling prey to cyber threats are reduced. On the contrary, the non-engagement of such adaptive cyber behaviours significantly compromises one's level of resilience on the electronic platform, which can be detrimental to the cyber health of computer users. It becomes imperative to explore this phenomenon of non-action and identify the 'what's and the 'why's of not adhering to cyber hygiene practices despite the obvious value of these behaviours.

[a]Such as: Singapore's First National Cybersecurity Awareness Campaign launched by CSA in early 2017 (CSA, 2017a); United Kingdom's GetSafeOnline Campaign (Get Safe Online, 2018) and Cyber Aware Campaign (Cyber Aware, n.d.); as well as Cyber Security Campaign in Hong Kong (Cyber Security Campaign, n.d.).

## 14.4 Awareness Does Not Equate to Increased Frequency of Cyber Hygiene Practices

It is vital that people do not just know about cyber hygiene practices but are also proactively undertaking such behaviours to protect themselves from cyber threats. The conventional understanding is that knowing how to and why one should engage in a certain behaviour should lead to the actual execution of the behaviour. However, awareness of cyber hygiene practices does not necessarily translate to its behavioural manifestation. People still fail to follow cyber hygiene practices despite becoming increasingly aware about cybersecurity issues [Schick, 2018].

### 14.4.1 *Understanding poor online behaviours based on insights from cyber hygiene surveys*

Surveys conducted on the frequency of cyber hygiene practices of individuals across the world reveal similar trends. In the United States, the Ponemon Institute conducted a US-based survey of consumers titled: "The Cyber Hygiene Index: Measuring the Riskiest States." The results found that Americans displayed risky cyber habits (e.g., not downloading and using anti-virus software) despite being aware that they should not do so [Moffitt, 2018]. Organisational employees were also engaging in risky behaviours (e.g., clicking on unverified links) despite fears of data breach, according to a survey by OpenVPN in America [Abel, 2018].

Crucially, findings from overseas surveys have been observed in Singapore as well. A 2017 cyber hygiene survey[b] (CHS) conducted by the Home Team Behavioural Sciences Centre (HTBSC) revealed that while individuals tend to know why they should engage in good cyber habits and

[b] In collaboration with social engineering expert Dr. Arun Vishwanath from the Department of Communication in University at Buffalo, HTBSC conducted a survey that assesses for various facets of Singaporeans' cyber hygiene. In particular, it assesses for: (1) awareness of cyber hygiene practices; (2) frequency of cyber hygiene practices (a total of 39 different practices were assessed); (3) risk beliefs; (4) phishing knowledge; and (4) reporting behaviour. A total of 404 responses were collected from the community (using convenience sampling method), with respondents from ages 16 to 67.

how to practice them, they are not really doing it.[c] With that, the 2017 CHS had also identified several positive cyber hygiene behaviours that respondents reflected both high awareness of, and displayed high frequency in adopting. These include checking the header of an incoming or new email to filter spam from genuine emails; not clicking on hyperlinks from unknown senders; as well as using ad-block tools to block pop-ups.

Knowledge should stimulate constructive cyber behaviours, just as how education is needed to inform people the 'why's and 'how's of doing something. Yet, the discrepancy between awareness and frequency of cyber hygiene practices is worrying, because it highlights underlying concerns that go beyond simply what education and information can do. The examination of motivations underlying this phenomenon of non-action is henceforth critical, since it may contribute to efforts to improve individuals' cyber hygiene levels.

## 14.5 There Are Many Psychological Reasons for Inaction of Cyber Hygiene Practices

The reasons for inaction of cyber hygiene practices cannot be attributed to a single factor. From a psychological perspective, various reasons can be identified to explain why people are not taking proactive steps to protect themselves from cyber threats, despite knowing the importance of doing so.

### 14.5.1 *Lack of individual responsibility*

The sense of individual responsibility can determine whether people will take necessary steps to protect themselves from cyberattacks. Responsibility is found to be related to action [Coleman, 2012], where

[c] According to the CHS, some top problematic cyber hygiene practices (in which people know why and how they should do it), are (in no sequence): using incognito mode or private mode when routinely surfing the internet; create new/unique logins and passwords for all online sign-ins; using a VPN when in an open/public Wi-Fi network; clearing browser cache; creating complex logins and passwords, checking device to ensure that it has the latest OS, software update, or patch; clearing cookies on browser; changing default passwords on all internet-enabled devices; managing how browser stores passwords; storing logins and passwords on encrypted online password vaults; managing privacy settings on social media platforms; keep virus protection updated; and running a virus scan on any new USB or external storage device.

people with a low sense of individual responsibility tend to be less motivated to engage in safe cyber hygiene practices.

Additionally, personal responsibility can also be influenced by the notion of self-licensing, whereby "positive behaviour disinhibits people from performing negative behaviour" [Merritt, Effron, & Monin, 2010, p. 350]. In the context of cyber hygiene practices, self-licensing could mean that people do not engage sufficiently in a variety of good practices, simply because they have implemented certain other practices. For instance, an individual may be less careful in terms of the websites that he or she visits, because he or she has activated the anti-virus programme on the computer.

In fact, cybersecurity research in organisational settings have demonstrated that people usually engage in these protective practices if relevant social others are doing it (e.g., peers, colleagues, direct superiors), or simply, if they are likely to get caught for not doing so [Herath & Rao, 2009]. A 2016 survey conducted by HTBSC on 'Perceptions of National Resilience'[d] further supported the notion of the lack of individual responsibility for dealing with cyberattacks. Far from seeing it as a responsibility that one should undertake, 64% of the respondents perceived that it is the government's duty instead to protect them from a cybersecurity crisis.

### 14.5.2 *Low cyber risk perception and complacency*

Optimism bias refers to the tendency for humans to underestimate the likelihood of negative events happening onto them [Sharot, 2011]. In the context of cyber threats, it can result in a false sense of security, since individuals are likely to perceive themselves to be 'invulnerable' to attacks [Shah, 2014]. Low cyber risk perception and complacency are likely to cause people to disregard cues of danger, behave recklessly online, and not take precautions to protect themselves from cyber threats.

The lack of experiences may further render individuals incognisant of how real the threat and consequences of cyberattacks are. Indeed, the absence of 'painful lessons' makes it difficult for one to justify why they

[d]A segment of the 'Perceptions of National Resilience' survey assesses individuals' perceptions of how likely various crisis scenarios will happen in Singapore, and to what extent are individuals, community, and government responsible for responding to crisis scenarios such as cybersecurity crisis. A representative total of 3,000 Singaporeans' responses was collected.

should learn, engage in efforts, as well as mobilise valuable resources to build some degree of cyber defence [Siau, 2017]. Furthermore, people may not necessarily engage in appropriate cyber hygiene behaviours, unless they deem that cyber threats might occur to them, and that these threats target their personal interests [Lee, Larose, & Rifon, 2008]. As such, this could be a potential reason why simply educating people on cyber threats and protection is insufficient.

### 14.5.3 *Reduced usability security as trade off*

The implementation of cyber security measures will increase one's protection against cyber threats. However, it may also reduce the ease of use of technology (i.e., speed, convenience) and one's work productivity and efficiency on technological devices [Bruzzese, 2015]. This is also known as the lack of usability security. When people realise that they may experience reduced speed and convenience, they are less likely to engage in good cyber hygiene practices even if it means acquiring tougher security levels.

For instance, on an individual level, while a newer mobile phone operating system (OS) patch provides greater and more effective security measures, people may hesitate to update their OS if it slows down the processing speed of their phone. Additionally, if the process of updating the software of a computer is a slow endeavour**,** for instance, on an organisational level, if many computers require updating, substantial opportunity costs between downtime and financial profits could result for said organisation. Consequently, the company may be tempted to forgo the benefits of the latest computer software update, and instead dedicate time to pursue financial earnings. Thus, whether or not individuals engage in cyber hygiene practices depend on the opportunity costs and their attitude towards it. If the consequences are positive, a favourable attitude is likely to transpire, thereby increasing the likelihood that the cyber hygiene practice will materialise.

### 14.5.4 *Difficulty in learning and implementing good cyber hygiene habits*

The more technical a cyber hygiene practice is for one to perform, the less 'common sense' the practice becomes [Ng, 2017]. The harsh truth is that if a cyber hygiene practice is perceived to be difficult to do, an individual naturally will be less likely to adopt that practice. For instance, checking if

the secure sockets layer (SSL) is present is important before one conducts a financial transaction via the internet. SSL refers to the lock icon beside the web address at the address bar of an internet browser, which is an indication as to whether a website is securely connected to the web server or not [GlobalSign, 2018]. However, it can be a challenging thing to do if people are not aware of what exactly is SSL, or the steps needed to check the quality of the SSL certificate. Consequently, people may become less likely to check if the SSL is present and/or the quality of the SSL certificate.

On the opposite spectrum, cyber hygiene practices that are easy to execute and commit to are more likely to see a higher frequency of behavioural manifestation of such practices. This could be a potential reason for the high awareness and frequency of several behaviours, including the checking of email header and using ad-block tools to block pop-ups, since they do not require significant conscious effort to do so. In other words, the individual does not have to remember that he or she has to do something (i.e., the cyber hygiene practice) as well how to do it.

### 14.5.5 *Lack of others who are also engaging in cyber hygiene practices*

Normative beliefs focus on whether a person feels pressured to do something or not, based on what he or she observes others to be doing [Ajzen, 1991, 2002]. In other words, individuals are likely to feel the need to engage in certain cyber practices simply because others are doing it. According to Herath and Rao [2009], the expectations and behaviours of relevant others, including superiors and peers, are significant influences on individuals' cyber security behaviours.

The awareness of why people engage or do not engage in certain cyber hygiene behaviours will be useful to inform and shape cybersecurity measures. People exhibit inaction for distinctive reasons, which perhaps then, lead to the requirement of a myriad of human-centric solutions—on top of technical security measures—to improve human cyber behaviours and reduce one's susceptibility to cyber threats.

## 14.6 Implications for Designing a Human-Centric Approach to Combat Cyber Threats

The reasons as to why people do or not do certain cyber hygiene practices differ from person to person, and even context to context. The development

of measures to improve problematic cyber behaviours should therefore be highly targeted and relevant, depending on the individuals' needs and the current climate of cyber threats. The following are some implications for designing human-centric measures against cyber threats.

### 14.6.1 *Public awareness efforts on cyber threats and cyber hygiene practices must continue*

Education is critical to help people understand why it is necessary for them to engage in good cyber habits. Consequently, it should prompt people to act to protect themselves from cyber threats, thereby reducing the inconsistency between people's awareness of cyber hygiene practices and how often they perform them. More importantly, besides informing people about the current climate of cyber threats, educational efforts should also cover (1) how cyber perpetrators target victims (i.e., exploiting the lack of good cyber behaviours), as well as the common methods they use to execute their attacks (i.e., social engineering); and (2) reasons that reduces the likelihood of practising these proper cyber practices so as to draw attention to this area of concern.

Regular mandatory cyber hygiene or cybersecurity programmes could be offered to people, such as the employees of an organisation, or students from an educational institution. In particular for firms that possess sensitive data, employees could also be informed and trained on identifying what these data are, and how it should be protected. In addition, they should be taught to understand the dire consequences if such sensitive information is compromised. These trainings can serve as an effective platform to instil knowledge, address any misconceptions on cybersecurity issues, or for individuals to share and learn from one another on their experiences, if any, in dealing with cyber threats. Participants do not have to be an actual victim before they can learn. The sharing and illustration of authentic experiences can aid in helping participants internalise the seriousness of cyber threats, whereby the fidelity of the victim's voice reinforces how real the threat and consequences are.

Unfortunately, there is also the need to revise educational messages occasionally, in order to recapture people's attention over time [Belch & Belch, 2009]. An underlying reason for doing so arises from the concern that constant exposure to the same messages can ultimately result in

individuals being desensitised—or numbed—towards these messages [So, Kim, & Cohen, 2017]. In other words, rephrasing or changing how outreach messages are delivered to people can facilitate for the attention-grabbing and retention of critical cyber information.

The regular revision of outreach messages will also be essential in ensuring that the public is receiving up-to-date information about the changing climate of cybersecurity. This is necessary, because the nature of cyberspace as well as cyberattacks—or their modus operandi—are constantly evolving [Daniel, 2017; Eidam, 2016].

### 14.6.2 *Tailor outreach messages*

Various sources of information can be condensed into digestible bite-sized knowledge and contribute to educational material. Local and international cyber threats can be studied holistically to identify new or emerging trends. Primary data may also be gathered to contribute to teaching resources, including interviews with incarcerated attackers to understand common modus operandi and the human vulnerabilities they target to facilitate their attacks.

Outreach messages ideally should be tailored to different societal groups, according to various demographic markers such as age—i.e., the types of cyber threats affecting the young and the old may differ, dictating the need to introduce tailored measures. Furthermore, targeted messages may help particular groups to understand and better relate to the current climate of cybersecurity using appropriate case examples and interventions. When the target audience understands why cybersecurity is an issue for them, it becomes easier for them to change their attitudes. Constant reminders are necessary as well, since people are likely to forget and become complacent over time. Different modes of transmitting these messages can be utilised. For people who do not have sufficient time to read their emails or cyber security advisories, sending concise email reminders on the importance of practicing good cyber habits is a feasible alternative.

There is also the need to revise educational messages occasionally, in order to recapture people's attention over time [Belch & Belch, 2009]. An underlying reason for doing so arises from the concern that constant exposure to the same messages can ultimately result in individuals being desensitised towards these messages [So *et al*., 2017].

### 14.6.3 *The need to nudge humans to engage in cyber hygiene practices*

Education by itself is clearly not sufficient. This was seen in the 2017 CHS results, in which people are still not engaging in important cyber actions despite knowing why or how they should do it. This highlighted underlying issues that goes beyond what education can do. The nudge theory can be employed to complement outreach efforts. A nudge prompts people to take action by subtly making the decision-making process easier [Chu, 2017]. The decisions that an individual have to make, or the thought processes that they have to undergo first, can be complicated enough such that it prevents or slows down the occurrence of engaging in good cyber behaviours altogether. Automation, for instance, is a nudging method whereby people are automatically required to do something unless they opt-out of it (if there is an opportunity to do so). In the context of cyber hygiene practices, this could mean setting various practices as a default feature or behaviour (i.e., natural 'opt-in') that people have to engage in when they use the computer or internet. As a result, it removes the need for people to proactively undertake constructive cyber practices. Higher participation and compliance to regulations are consequently more likely to follow as well [Blau, 2017]. Automation can also make certain practices less difficult or tedious to do (e.g., reduces the amount of time, effort, and knowledge needed to perform it), thereby increasing the likelihood that people will eventually engage in these protective measures.

### 14.6.4 *Conduct red-teaming exercises*

The objective of red-teaming exercises is to "obtain a realistic level of risk and vulnerabilities against your technology, people and physical/facilities" [RedTeam Security Consulting, 2016, p. 1]. One method to do so is to invite white hats to hack into these systems, thereby checking the resiliency of an organisation's computer and network systems. In other words, a cyberattack will be purposefully—with no malicious intention—conducted without employees' knowledge of the exercise, to reveal the ways in which cyber perpetrators can exploit to enter the organisation's computer and network systems. This includes identifying both human and technical vulnerabilities.

The benefit of red-teaming also lies in its capacity to create a level of psychological fidelity, allowing people to experience the effects of a real

cyberattack but under safe circumstances with minimal repercussions. According to Boud, Cohen, and Walker [1993], such experiences can help to accentuate the criticality of cyber issues within the organisation. With red-teaming exercises, it is with the expectation and hope that people will be more careful and thus practice good cyber hygiene practices in the future, after experiencing the negative emotions and consequences of a cyberattack.

Nonetheless, there is a need to balance between the endeavour of identifying the vulnerabilities of one's online security system and the potential negative emotional exerted on the "victims" as a result of such exercises. Should red-teaming exercises be conducted, there is a need to provide debriefs to these individuals to help manage their reactions.

### 14.6.5 *Regulations to ensure compliance to cyber hygiene practices*

Certain cyber threats lay dormant and are not obvious unless a computer or network system is consciously scanned for compromise. Unfortunately, the notion of reviewing these systems on a regular basis may not be welcomed, particularly when the process is time-consuming and effortful. Regulations to drive for consistent audits should hence be implemented, such as the need to ensure that employees conduct regular anti-virus scans to eliminate any dormant malicious software on their work computers. These regulations may also include policies to ensure that people are mandated to engage in other necessary cyber hygiene practices. For instance, the Monetary Authority of Singapore (MAS)—being the first financial authority in the world to do so—is implementing cyber hygiene rules that financial institutions in Singapore must adhere to, starting August 2020, as a means to heighten the industry's resilience against cyber risks ["New cyber hygiene rules," 2019].

In addition, detection mechanisms to detect lack of compliance could be developed and executed alongside these regulations. As expressed by Herath and Rao [2009], people will observe rules simply because of the fear of being discovered to be not observing them. To ensure its effectiveness, such detection mechanisms initiatives should be clearly communicated to the employees in advance, and constantly reinforced.

A subtler option to get people to start complying with cyber hygiene practices is ensuring that superiors or leaders in an organisation are doing

so as well, since people are likely to model after others who are deemed to be important and significant characters within the organisation. In the long run, this may create a culture in the organisation that engaging in good cyber hygiene is common and normal, and that compliance to cybersecurity protocols is indeed necessary.

## 14.7 Conclusion

The common denominator across most cyber hygiene guidelines revolves around the implementation of cybersecurity 'best practices' to protect and maintain one's online 'safety' and 'health'. It focuses on individual responsibility to perform the identified best practices, rather than depending on technical protection measures against cyber threats. It is also not about a particular online behaviour that a person exhibits, but rather the totality of the individual's positive and negative cyber behaviours.

Compounded with the rising sophistication, technical capabilities, and resources of the perpetrators in the cyber arena, the challenges of developing good cyber practices in individuals inevitably become more demanding. As such, cyber practices need to be differentiated and contextualised accordingly to the environment and circumstance that they should be applied in—such as in the setting of a home or workplace. Moreover, since existing complexities are further compounded by the end-user interactions with the online platform, this predates the need for different cyber practices for different user interactions.

## 14.8 Acknowledgement

The views expressed in this chapter are the author's alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 14.9 References

Abel, R. (2018, June 18). Despite advancements, employees still practice bad cyber-hygiene, study. *SC Media.* Retrieved from https://www.scmagazineuk.com/despite-advancements-employees-practice-bad-cyber-hygiene-study/article/1486713

Accenture and Ponemon Institute. (2017). *2017 cost of cyber crime study: Insights on the security investments that make a difference*. Retrieved from https://www.accenture.com/t20170926T072837Z__w__/us-en/_acnmedia/PDF-61/Accenture-2017-CostCyberCrimeStudy.pdf

Aldorisio, J. (2018, March 26). What is cyber hygiene? A definition of cyber hygiene, benefits, best practices, and more. *Digital Guardian.* Retrieved from https://digitalguardian.com/blog/what-cyber-hygiene-definition-cyber-hygiene-benefits-best-practices-and-more

Ashiq, J. (2015, April 30). The Importance of cyber hygiene in cyberspace. *Infosec Institute.* Retrieved from http://resources.infosecinstitute.com/the-importance-of-cyber-hygiene-in-cyberspace/#gref

Belch, G. E., & Belch, M. A. (2009). Source, message, and channel factors. In *Advertising & promotion: An integrated marketing communications perspective* (8th ed., pp. 174–205). Boston: McGraw-Hill Irwin.

Bennett, M. (2017). Building a digital security army. *The Telegraph.* Retrieved from https://www.telegraph.co.uk/business/digital-security/human-behaviour-in-digital-security/

Blau, A. (2017, December 11). Better cybersecurity starts with fixing your employees' bad habits. *Harvard Business Review.* Retrieved from https://hbr.org/2017/12/better-cybersecurity-starts-with-fixing-your-employees-bad-habits

Boud, D., Cohen, R., & Walker, D. (1993). *Using experience for learning*. Bristol, PA: The Editors and Contributors.

Choo, B., Dillon, L., Neo, L. S., Ong, G., Tan, E., & Khader, M. (2016). *Social engineering: Using psychology to exploit bugs in the human operation system* (HTBSC Research Report No.: 01/2016). Singapore: Ministry of Home Affairs, Home Team Behavioural Sciences Centre.

Chu, B. (2017, October 9). What is 'nudge theory' and why should we care? Explaining Richard Thaler's Nobel economics prize-winning concept. *Independent.* Retrieved from https://www.independent.co.uk/news/business/analysis-and-features/nudge-theory-richard-thaler-meaning-explanation-what-is-it-nobel-economics-prize-winner-2017-a7990461.html

Coleman, J. (2012, August 30). Take ownership of your actions by taking responsibility. *Harvard Business Review.* Retrieved from https://hbr.org/2012/08/take-ownership-of-your-actions

Conteh, N. Y., & Schmick, P. J. (2016). Cybersecurity: Risks, vulnerabilities and countermeasures to prevent social engineering attacks. *International Journal of Advanced Computer Research*, *6*(23), pp. 31–38.

CSA. (2017a, February 10). *CSA launches First National Cybersecurity Awareness Campaign* [PDF document]. Retrieved from https://www.gov.sg/~/sgpcmedia/media_releases/csa/press_release/P-20170210-1/attachment/CSA%20Launches%20First%20National%20Cybersecurity%20Awareness%20Campaign_10%20Feb%202017.pdf

CSA. (2017b, February 15). *CSA releases key findings from first Cybersecurity Public Awareness Survey* [PDF document]. Retrieved from http://www.nas.gov.sg/archivesonline/data/pdfdoc/20170215003/140217_CSA%20Releases%20key%20findings%20from%20first%20cybersecurity%20public%20awareness%20survey.pdf

Curtin Singapore's website defaced by hackers claiming to represent ISIS. (2015, March 10). *Today.* Retrieved from https://www.todayonline.com/singapore/curtin-singapores-website-defaced-hackers-claiming-represent-isis

Cyber Aware. (n.d.). Retrieved from https://www.cyberaware.gov.uk

Cyber Security Campaign. (n.d.). Retrieved from https://www.cybersecuritycampaign.com.hk/index-en.html#

Daniel, M. (2017, May 22). Why is cybersecurity so hard? *Harvard Business Review.* Retrieved from https://hbr.org/2017/05/why-is-cybersecurity-so-hard

Drzik, J. (2018, January 17). Cyber risk is a growing challenge. So how can we prepare? *World Economic Forum.* Retrieved from https://www.weforum.org/agenda/2018/01/our-exposure-to-cyberattacks-is-growing-we-need-to-become-cyber-risk-ready

Eidam, E. (2016, May 24). Cybersecurity is an ever-changing landscape. *government technology.* Retrieved from https://www.govtech.com/security/Cybersecurity-Is-an-Ever-Changing-Landscape.html

Fallows, J. (2011, March 24). Cyber-security can't ignore human behaviour. *The Atlantic.* Retrieved from https://www.theatlantic.com/technology/archive/2011/03/cyber-security-cant-ignore-human-behavior/72826/

Foo, S., & Jayakumar, S. (2018, January 26). Cyber threats: 2018 and beyond. *The Straits Times: Opinion.* Retrieved from https://www.straitstimes.com/opinion/cyber-threats-2018-and-beyond

Get Safe Online. (2018). Retrieved from https://www.getsafeonline.org

GlobalSign. (2018). *What is SSL?* Retrieved from https://www.globalsign.com/en-sg/ssl-information-center/what-is-ssl/

Goldman, D. (2013, January 31). Your antivirus software probably won't prevent a cyberattack. *CNN tech.* Retrieved from https://money.cnn.com/2013/01/31/technology/security/antivirus/index.html

Graham, L. (2017, September 20). The number of devastating cyberattacks is surging—and it's likely to get much worse. *CNBC.* Retrieved from https://www.cnbc.com/2017/09/20/cyberattacks-are-surging-and-more-data-records-are-stolen.html

Greig, J. (2018, April 19). Why human vulnerabilities are more dangerous to your business than software flaws. *TechRepublic.* https://www.techrepublic.com/article/why-human-vulnerabilities-are-more-dangerous-to-your-business-than-software-flaws/

Hadlington, L. (2017). Human factors in cybersecurity: Examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon*, *3*(7), e00346.

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: A framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), pp. 106–125.

Iswaran, S. (2018, August 6). Statement by Mr. S. Iswaran, Minister-in-Charge of Cybersecurity, on the cyber-attack on SingHealth's IT system, during Parliamentary Sitting, 6 August 2018. *Ministry of Communications and Information.* Retrieved from https://www.mci.gov.sg/pressroom/news-and-stories/pressroom/2018/8/statement-by-mr-s-iswaran-on--cyber-attack-on-singhealth-it-system-during-parl-sitting-on-6-aug-2018

Kelly, R. (2017, March 3). Almost 90% of cyber attacks are caused by human error or behaviour. *Chief Executive.* Retrieved from https://chiefexecutive.net/almost-90-cyber-attacks-caused-human-error-behavior/

Kwang, K. (2018, July 20). Singapore health system hit by 'most serious breach of personal data' in cyberattack; PM Lee's data targeted. *Channel NewsAsia.* Retrieved from https://www.channelnewsasia.com/news/singapore/sing-health-health-system-hit-serious-cyberattack-pm-lee-target-10548318

Lee, D., Larose, R., & Rifon, N. (2008). Keeping our network safe: A model of online protection behaviour. *Behaviour & Information Technology*, *27*(5), pp. 445–454.

Leech, M. (2017, September 21). Data breach statistics 2017: First half results are in. *Gemalto.* Retrieved from https://blog.gemalto.com/security/2017/09/21/new-breach-level-index-findings-for-first-half-of-2017/

Lewis, J. A. (2002). *Assessing the risks of cyber terrorism, cyber war and other cyber threats*. Washington, DC: Center for Strategic and International Studies.

Lim, J. (2013, October 30). Ang Mo Kio Town Council website hacked. *The Straits Times.* Retrieved from https://www.straitstimes.com/singapore/courts-crime/ang-mo-kio-town-council-website-hacked

Loke, K. F. (2017, February 28). MINDEF Internet system breached; data stolen from national servicemen, employees. *Channel NewsAsia.* Retrieved from http://www.channelnewsasia.com/news/singapore/mindef-internet-system-breached-data-stolen-from-national-servic-7617146

Lord, N. (2017, July 27). What is a phishing attack? Defining and identifying different types of phishing attacks. *Digital Guardian.* Retrieved from https://digitalguardian.com/blog/what-phishing-attack-defining-and-identifying-different-types-phishing-attacks

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*. *4*(5), pp. 344–357.

Ministry of Health. (2019, January 28). *Unauthorised possession and disclosure of information from HIV registry.* Retrieved from https://www.moh.gov.sg/news-highlights/details/unauthorised-possession-and-disclosure-of-information-from-hiv-registry

Moffitt, T. (2018, June 5). American cybercrime: The riskiest states in 2018. *Webroot: Smarter Cybersecurity.* Retrieved from https://www.webroot.com/blog/2018/06/05/2018-riskiest-states-for-cybercrime-in-america/

Mokhtar, F. (2018, June 19). Cyber threats in Singapore go up; phishing attacks see biggest jump. *Today.* Retrieved from https://www.todayonline.com/singapore/cyber-threats-singapore-go-phishing-attacks-see-biggest-jump

New cyber hygiene rules for financial services, e-payment firms to kick in next August: MAS. (2019, August 6). *The Business Times.* Retrieved from https://www.businesstimes.com.sg/banking-finance/new-cyber-hygiene-rules-for-financial-services-e-payment-firms-to-kick-in-next

No Internet access for public officers' work computers by next June. (2016, June 8). *Channel NewsAsia.* Retrieved from https://www.channelnewsasia.com/news/singapore/no-internet-access-for-public-officers-work-computers-by-next-ju-7961140

Osborne, C. (2019, December 12). These are the worst hacks, cyberattacks, and data breaches of 2019. *ZDNet.* Retrieved from https://www.zdnet.com/article/these-are-the-worst-hacks-cyberattacks-and-data-breaches-of-2019/

PwC. (2016a). Adjusting the lens on economic crime. *Global Economic Crime Survey 2016* [PDF document]. Retrieved from https://www.pwc.com/gx/en/economic-crime-survey/pdf/GlobalEconomicCrimeSurvey2016.pdf

PwC. (2016b). *PwC's 2016 Global Economic Crime Survey—Singapore* [PDF document]. Retrieved from https://www.pwc.com/sg/en/consulting/assets/economic-crime-survey/economic_crime_survey_2016_singapore.pdf

PwC. (2017). *Global State of Information Security Survey 2017: Singapore highlights* [PDF document]. Retrieved from https://www.pwc.com/sg/en/risk-assurance/assets/gsiss/global-state-of-information-security-survey-2017-sg.pdf

RedTeam Security Consulting. (2016). *Full Force Red Team.* Retrieved from https://www.redteamsecure.com/red-teaming/

Schick, S. (2018, June 13). Poor password practices put corporate cybersecurity at risk. *Security Intelligence.* Retrieved from https://securityintelligence.com/news/poor-password-practices-put-corporate-cybersecurity-at-risk/

Shiao, V. (2017, August 1). A third of Singapore SMEs hit by ransomware last year: study. *The Business Times.* Retrieved from https://www.businesstimes.com.sg/technology/a-third-of-singapore-smes-hit-by-ransomware-last-year-study

Singtel. (2018, May). *Managing Cyber Security Incidents Before They Become Crises*. Retrieved from https://www.singtel.com/business/singtel-global-services/content/managing-cyber-security-incidents-before-they-become-crises

So, J., Kim, S., & Cohen, H. (2017). Message fatigue: Conceptual definition, operationalization, and correlates. *Communication Monographs*, *84*(1), pp. 5–29.

Strategy for a Technology-driven Future. (2017, November 3). *Infocomm Media Development Authority.* Retrieved from https://www.imda.gov.sg/infocomm-and-media-news/buzz-central/2016/6/strategy-for-a-technology-driven-future

Symantec Corporation. (2017). Good cyber hygiene. *Norton by Symantec.* Retrieved from https://us.norton.com/internetsecurity-how-to-good-cyber-hygiene.html

Tasman-Jones, J. (2016, March 31). Human behaviour still biggest cause of cybercrime. *Fund Strategy.* Retrieved from https://www.fundstrategy.co.uk/human-behaviour-still-biggest-cause-of-cybercrime/

Tham, I. (2017a, May 12). Hackers broke into NUS, NTU networks in search of government, research data. *The Straits Times.* Retrieved from http://www.straitstimes.com/singapore/hackers-broke-into-nus-ntu-networks-in-search-of-government-research-data

Tham, I. (2017b, May 21). Cyber hackers and digital defences: Gone phishing … So, everyone, on guard. *The Straits Times.* Retrieved from http://www.straitstimes.com/tech/gone-phishing-so-everyone-on-guard

Tham, I. (2017c, September 7). AXA data breach affects 5,400 Singapore customers. *The Straits Times.* Retrieved from http://www.straitstimes.com/singapore/axa-data-breach-affects-5400-singapore-customers

Tham, I. (2018). Personal info of 1.5m SingHealth patients, including PM Lee, stolen in Singapore's worst cyber attack. *The Straits Times.* Retrieved from https://www.straitstimes.com/singapore/personal-info-of-15m-singhealth-patients-including-pm-lee-stolen-in-singapores-most

Vishwanath, A. (2016, May 5). Cybersecurity's weakest link: Humans. *The Conversation.* Retrieved from https://theconversation.com/cybersecuritys-weakest-link-humans-57455

World Economic Forum. (2018). *The Global Risks Report 2018* [PDF document]. Retrieved from http://www3.weforum.org/docs/WEF_GRR18_Report.pdf

This page intentionally left blank

# Section F

# The Future of Cybersecurity

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0015

# Chapter 15

# Smart Homes: Where Rogue AI and Robots could Impair Security

Muhammad Faizal B Abdul Rahman

*Centre of Excellence for National Security,*
*S. Rajaratnam School of International Studies,*
*Nanyang Technological University*

*ismfaizal@ntu.edu.sg*

> "All devices have their dangers. The discovery of speech introduced communication—and lies. The discovery of fire introduced cooking—and arson. The discovery of the compass improved navigation—and destroyed civilisations in Mexico and Peru."
>
> — Isaac Asimov

## 15.1 Introduction

Since time immemorial, people have imagined intelligent machines that could assist them in performing tasks more efficiently. In his introductory essay for the book *Robot Visions*, Isaac Asimov wrote that such machines would be "capable of performing tasks of a kind that are too complex for any living mind other than that of a man" [Asimov & McQuarrie, 1990, p. 2]. He cited stories from mythology in which Hephaestus, the Greek God of Forge, had gold-based helpers who were capable of thinking, communicating, and performing manual labour. Talos, a bronze giant, patrolled the coastlines of Crete to defend the island against invaders.

These stories bear similarities to the robots that would pervade the Fourth Industrial Revolution (4IR) era.

In the foreseeable future, one space and context where people are likely to interact with artificial intelligence (AI) and robots is their residence—smart homes. According to market research, the global home automation market is expected to reach US$114 billion by 2025, a marked increase from US$45.8 billion in 2017 [Fortune, 2019]. Advances in AI, cloud computing, internet-connected devices (IoT), robotics, virtual assistants, and 5G technology are driving seamless automation of homes closer to reality [Chen, 2020)]. More importantly, programmers designed AI to provide six core capabilities: (1) activity recognition; (2) image recognition; (3) voice recognition; (4) data processing; (5) prediction making; and (6) decision-making. These capabilities could support the various functions—such as energy management, healthcare and security—of smart homes [Guo *et al*., 2019]. Additionally, the report "Artificial Intelligence and Life in 2030" by Stanford University predicts that the integration of AI in robotics would hasten the adoption of home robots by 2030 [Bharadwaj, 2019b]. Robots could perform three home-related essential and manual tasks: (1) cleaning; (2) entertainment; and (3) security and surveillance [Bharadwaj, 2019b]. As smart home technologies become increasingly available in the market, the cost of their adoption would decrease. Besides the uber-rich and technologically-savvy, more average home owners could embrace home automation. Eventually, smart homes would become integral to the daily lives of people.

Asimov, however, cautioned that technology has its risks too. Despite the growing importance of cybersecurity amid the increase in numbers and functions of smart devices at homes, incidents of data breaches, distributed denial of service (DDoS), and ransomware attacks continue unabated. For example, in 2016, the Mirai botnet launched widespread DDoS attacks by controlling thousands of home IoT devices, using them as zombie machines to disrupt major websites including Twitter, Cable News Network (CNN), PayPal, and Amazon.com [Bursztein, 2017]. Looking ahead, the report "The New Norm: Trend Micro Security Predictions 2020" predicted that threat actors would increasingly explore the use of AI to hack smart devices for extortion and espionage. The decade leading up to 2030 could see an expansion of cyberattack surface as a result of the growth of home automation, especially in highly digitalised societies and developed countries.

This chapter will first discuss key points on cyberattacks involving AI and robots in smart homes. Next, this chapter will discuss approaches that cyber defenders should consider in protecting highly digitalised societies and developed countries where smart homes could be the norm in the future.

## 15.2 Cyber Threats Involving Smart Homes

As with any other technology, smart homes have their own set of flaws. Foremost, engineers, and programmers could never build perfect defences against known and potential threat actors. Software codes and algorithms could have errors. The lines of codes in all devices increase as they become smarter and as more devices work together to constitute a smart home. This technical aspect of the smart home increases the risk of exploitable flaws. Connectivity further increases the exposure of these flaws to hostile online interactions with threat actors [Linday *et al*., 2016]. The following are five key points that could define cyber threats involving smart homes.

### 15.2.1 *Key Point #1: Why fear machine havoc in smart homes?*

The first key point is the notion that cyber threats involving smart homes constitute a phenomenon that is larger than data insecurity or service disruptions. From a philosophical perspective, a compromised smart home is the materialisation of fear. As Asimov has written, this is the fear of the potential harm that out-of-control machines could inflict on people. Like fire, smart homes that go rogue or become the tool of threat actors could create chaos instead of convenience. For example, a scene in episode one of season two of the 2016 sci-fi drama series *Mr. Robot* illustrated how threat actors could make a smart home torment its occupant in an act of hacktivism [Epstein, 2016].

The fear of machines going out of control increases as they become more intelligent and require less human control [Asimov & McQuarrie, 1990]. For example, the 1977 sci-fi horror movie *The Demon Seed* illustrated how an AI—after absorbing massive data—had gone rogue in its creator's home and terrorised his wife in an act of rebellion [Epstein, 2016]. Although these movies featured compromised smart homes in

fictional settings, they nonetheless highlighted two feared outcomes. Firstly, that of threat actors making smart homes' AI go rogue by feeding it with corrupt data, making changes, and introducing biases to its algorithms. Secondly, that of threat actors hijacking smart homes to target people in acts of cybercrime, protest, or cyberwar. In 2019, a survey by the Oxford University's Centre for Governance of AI found that many Americans do fear the risks—including possible human extinction—that could manifest years after machines attain high-level intelligence [Zhang & Dafoe, 2019].

The present pace of AI development and society's growing dependence on robots and cyberspace have made the of threat actors hijacking AI and robots in smart homes increasingly conceivable. For smart homes operated by AI and robots, a slight human oversight could result in occupants becoming belatedly aware of the security breaches (i.e., only when the threats materialise). Additionally, connectivity could turn an employee's smart home or a home office into a launching point to attack supply chains and corporate networks [Trend, 2019]. Smart homes could potentially evolve into a national security issue, given the far-reaching implications when compromised.

Next, threat actors who seek to undermine national security could perceive smart homes as attractive targets in the future. In societies where housing is a critical national policy, the security of smart homes could become a political issue that influences the social contract between the people and the government. In the international space, dominance in smart home technologies and market size could become a subset in the arena of geopolitical contestation between rival countries. The next four points will delve into this issue of national security.

### 15.2.2 *Key Point #2: AI and robots as agents of harm*

The second key point is the risk of threat actors hijacking AI-powered devices and robots in smart homes to harm the occupants. In a home setting, the constant human–machine interaction between the occupants and the AI-powered devices and robots could create an environment of trust over time, whereby the occupants would trust these intelligent machines to fulfil the functions of a home. Consequently, children may regard these machines—such as Alexa—that fulfil social needs as part of the family [Manku, 2018]. Threat actors could exploit both the proximity between these machines and the occupants, as well as the environment of

trust to cause harm. By hacking intelligent machines, threat actors could turn them into agents of harm.

Research has suggested some methods that could make intelligent machines go rogue. Firstly, it may be possible to manipulate intelligent machines by tampering with the data or the physical environment that the AI in the machines process and observe to make decisions. This manipulation entails fooling intelligent machines to see something that is not there, ignoring what is there, and misclassifying objects and actions. Of course, threat actors would need knowledge on the AI's machine learning model to examine for weaknesses and run simulations to determine what changes to the data and physical environment would make manipulation successful [Kobie, 2019]. Secondly, threat actors could weaponise AI (i.e., adversarial AI) that adapts itself as it learns about its target's digital environment to compromise intelligent machines while evading detection. Traditional cybersecurity tools may be less effective against adversarial AI, as they are generally programmed to detect known and predictable attacks. Experiments by cybersecurity experts have demonstrated that adversarial AI could evade anti-malware tools [Kharkar *et al*., 2017]. Adversarial AI could therefore become powerful tools for threat actors who seek to outwit cyber defenders [Dixon & Eagan, 2019].

Intelligent machines that turn into agents of harm could act in various ways based on the intent of the threat actors. The example of hackable hotel robots in Japan demonstrated how intelligent machines could spy for threat actors and conduct illegal surveillance on occupants of smart homes [Linder, 2019]. The information that these intelligent machines collect, which could threaten privacy, could be channelled for mischievous, criminal, or political purposes. Cybersecurity experts have also demonstrated how they could hack and turn home robots such as Alpha-2 into a weapon [Vincent, 2017]. The fact that these experts could make Alpha-2 attack a tomato with a screwdriver foreshadows a disturbing future scenario in which threat actors could commandeer home robots to attack or threaten its occupants.

Compromised intelligent machines could also inflict more harm than what the threat actors had intended and produce unintended consequences. For example, the WannaCry ransomware outbreak in 2017 that was attributed to North Korea affected thousands of IoT devices around the world. The rapid spread of the malware suggests that the threat actors may have lost control of it due to design flaws

[Newman, 2019]. Likewise, there is the risk of adversarial AI becoming unpredictable and spinning out of control when hacking smart homes. The late Stephen Hawking once wrote that AI could outsmart humans and become difficult to control in the long term [Hawking *et al*., 2014]. This risk could present itself amid concerns of indoor AI-controlled systems such as home robots being compromised and causing more harm than it would have been possible if under human control [Brundage *et al*., 2018].

### 15.2.3 *Key Point #3: Motivations of threat actors*

The third key point is the different motivations of threat actors in targeting smart homes. The act of hacking smart homes could be synonymous with the crimes of housebreaking or home invasion, albeit through digital means. Additionally, home offices and remote-working arrangements would blur the lines between home and work, thereby creating intersections between the cyber threats that smart homes and corporate workplaces encounter. The smart home and its occupants could be the primary target of threat actors, or the gateway for enterprise network attacks [Trend, 2019]. Research suggests four classifications of motivations that could explain the act of hacking smart homes: (1) curiosity; (2) personal gain; (3) terrorism; and (4) geopolitical agendas. These motivations may also be aligned with the nature of the threat actors, which will be discussed moving forward, and is summarised in Figure 15.1 [Bugeja *et al*., 2017].

Firstly, curiosity could drive threat actors to conduct illicit experiments by testing their hacking skills on smart homes and against cybersecurity measures. Such threat actors are hackers who experience boredom and crave for intellectual challenges. As societies become more digitalised and more people learn computer and coding skills, it is likely that more curious people would gain the capabilities to conduct hacking-related activities. In the future, the hacking of smart infrastructure, including smart homes, could be the manifestation of anti-social behaviour [Klein, 2019]. For example, affirmation culture could be the reason why Michael Reeves, a software developer and YouTuber, programmed a Roomba vacuum robot to spew expletives when it comes in contact with objects [Sholihyn, 2019]. This culture encourages individuals to conduct and showcase “cool” activities—which may not be socially acceptable—in order to acquire praises and popularity on

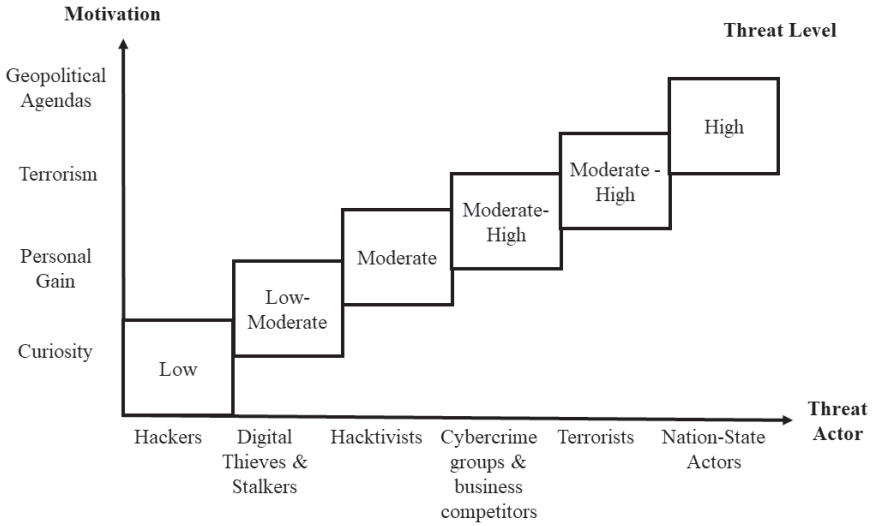



Figure 15.1. Primary motivations and threat levels of smart homes' cyber threat actors.
*Source*: Adapted from Bugeja *et al*., 2017.

social media [Wong, 2019]. Curious hackers pose the least threat as their actions stem from mischief rather than malice.

Secondly, personal gain (e.g., skills-building, financial benefits, and victimisation) is another motivation that could drive threat actors to hack smart homes. Their threat level is higher than curious hackers' as they would knowingly conduct malicious acts at the expense of the occupants. These threat actors hack to hone their skills in finding and exploiting cyber vulnerabilities. More worryingly, these hackers could "graduate" into digital thieves if they see profit in spying on smart homes and selling information to buyers in the digital black markets. They may use ransomware to demand money from occupants in exchange for not making the home robots malfunction and cause harm such as crashing or setting fire to the house [Ziska, 2018]. These threat actors could also be malicious individuals—such as cyber-bullies, stalkers, and predators—who view smart home devices as tools like social media that they can use to victimise and cause distress to people whom they hate or desire. Threat actors who demand radical sociopolitical changes would be hacktivists targeting the smart homes of influential individuals due to the latter's corporate or political activities that enrich only the rich and powerful. Smart home devices, if compromised, could also damage the reputation of their suppliers. The

threat level could rise if cybercriminal groups and unscrupulous businesses that seek to undermine competitors in the industry outsource their unlawful activities to or recruit these threat actors [Yusof, 2016].

Thirdly, terrorism—which may serve as an instrument in geopolitical agendas—is a motivation that could drive governments to prioritise the security of smart homes, much like that of national critical infrastructures. Terrorists are known to seek new opportunities and leverage new technologies in their tactics and goals to spread fear, communicate their demands, and compel sociopolitical changes [Veer, 2019]. Currently, the lack of resources and skill among terrorists and the lack of physical harm in known cyberattacks make terrorism the least likely motivation for hacking smart homes. Nonetheless, the growth in smart homes could create new opportunities for terrorism that are improbable today. What is certain today, however, is the fact that terrorists have demonstrated capabilities in using 4IR technologies to engage in asymmetric warfare in order to 'level the playing field'. For example, the Houthi insurgents in Yemen have deployed military-grade drones in its attack on oil facilities in Saudi Arabia [Altaher *et al*., 2019].

If terrorists do acquire the capabilities to hack smart homes, the possible sources of these capabilities are either cybercriminals or hostile nation-state actors who make use of terrorist groups as proxies in geopolitical conflicts. Ultimately, the terrorists' consideration for cyberattacks on smart homes is contingent on whether such attacks are instrumental to their group's grand strategy. For example, cyberattacks on smart homes should generate psychological effects that are almost comparable to the impact of targeted killings or multiple attacks on civilians. Taking a leaf from the works of military strategist Colin Gray, technology is "a team player in the gestalt that is strategy" and "drives tactics, shapes operations and enables strategies" [Gray, 2013, p. 168].

Lastly, geopolitical agendas could drive nation-state actors to exploit cyberspace as a domain of warfare. For example, cyberattacks in 2008 disrupted communications infrastructure in Georgia weeks before the country faced an invasion by Russian forces. The US and Israel had reportedly deployed the Stuxnet computer virus in 2007 to disrupt operations at an Iranian nuclear research facility. Today, cyberattacks not only serve as precursors to kinetic attacks, they are increasingly used as a means of foreign intervention and conducting conflict below the threshold of conventional war. In the future, smart homes could become targets of cyberattacks, in addition to the targeted country's national critical

infrastructures. Unlike national critical infrastructures, however, smart devices in homes are more susceptible to cyberattacks as they have less authentication and encryption features due to their smaller sizes and design for ease of consumer use [Toon, 2019]. Homes that are rendered unsafe by cyberattacks would be disruptive to daily lives. Over time, such disruption could potentially result in law and order issues, and impact the country's social harmony.

Furthermore, the world is approaching the era of machine-driven warfare, whereby militaries would utilise AI and robots for myriad roles [Zachary, 2019]. Compromised AI-powered devices and robots in smart homes could complement AI-powered weapon systems and change how wars are fought. For example, a military adversary could use AI-powered aerial drones to attack a country's government facilities while concurrently exploiting compromised AI-powered devices and robots in smart homes to cause distress to the civilian population. Threat actors who serve geopolitical agendas pose the highest threat level given the greater sophistication of their motivation and tactics, and the potential impact on the targeted country's people.

### 15.2.4 *Key Point #4: Smart homes are the door to a country's centre of gravity*

The fourth key point relates to smart homes becoming the frontline in future geopolitical conflicts. Since smart homes are not considered national critical infrastructures, cyberattacks on them may not pose a direct threat to national security. However, such cyberattacks affect the occupants—people who constitute the political constituency and economic lifeblood of the country. Fundamentally, people are the social centre of gravity (COG) of the targeted country [Porter, 2018]. Cyberattacks on smart homes could undermine the source of national power (i.e., people) from which the country draws its strength and the will to fight. Cyber defenders should understand that while smart homes are currently an emerging trend, they could become socially ubiquitous in the future. Their ubiquity could make them attractive targets in future grey zone conflicts where the rules of engagement are ambiguous.

As smart homes are civilian targets, it is possible that cyberattacks against them could serve as a tool of intimidation or coercive diplomacy, whereby the aim is to impede specific policies of the targeted country that are disadvantageous to the geopolitical interests of the threat actor.

Such cyberattacks, if they cause limited harm, could be non-escalatory and advantageous for the threat actor, especially if the costs of military action are considerable.

However, such attacks could be escalatory if that is the intent, and if the threat actor is reckless or overestimates its ability to control the adversarial AI and compromised home robots [NSI, 2019]. Conflict escalation could also occur if there is a vast difference in threat perceptions. For example, the threat actor could consider the cyberattack on smart homes an acceptable behaviour because there is no violation of territorial sovereignty or direct acts of violence, but the targeted country could interpret it as an act of war [Heinl, 2016].

Depending on the targeted country's strength, cyberattacks on smart homes could create powerful psychological effects on the country's people. They may perceive that the threat actor has overwhelming power, and this could diminish their morale and confidence in the country's leaders and defences. This effect is similar to when strategic aerial bombing on population centres during World War II was conducted to break the people's will [Dermer, 2013].

Additionally, the ambiguous nature of cyberattacks enables the threat actor to interpret the principles of the Law of Armed Conflict to its advantage. For example, the blurring of lines between combatants and non-combatants may happen in cyberspace, and this creates the legal ambiguity for the threat actor to launch non-kinetic attacks on civilian targets.

This ambiguity is useful against a targeted country that depends significantly on conscripts (i.e., citizen soldiers) for military defence. If the magnitude of harm from cyberattacks on smart homes becomes unbearable, the people could lobby their government to surrender to the threat actor in exchange for peace [Dermer, 2013]. Ultimately, the threat actor could strive for strategic decisiveness in geopolitical conflicts by hacking smart homes to undermine the targeted country's social COG. The nature of cyberspace allows the threat actor to minimise any military confrontations, surmount "the tyranny of distance" and hit the targeted country where it hurts most.

### 15.2.5 *Key Point #5: Smart homes, human rights, and basic needs*

The fifth key point relates to the notion of how cyberattacks on smart homes could undermine national security by depriving the people in the

targeted country of their human rights and basic needs. In this regard, Article 3 of the Universal Declaration of Human Rights (UDHR) states that "everyone has the right to life, liberty and security of person"; and Article 25 states that "everyone has the right to a standard of living adequate for the health and well-being … including food … housing …" [Yacine, 2015, pp. 8 and 52]. In her speech on October 2019, the UN High Commissioner for Human Rights Michelle Bachelet explained the intersection of human rights and smart technologies. She said "… Neither can we afford to see cyberspace and artificial intelligence as an ungoverned or ungovernable space—a human rights black hole. The same rights exist online and offline" [UN, 2019]. Therefore, it would be rational to presume that smart homes, too, should facilitate the fulfilment of these human rights.

A profound relationship exists between human rights and people's basic needs. Research suggests that Maslow's hierarchy of needs provides a framework to "define human rights in terms of human needs and linking them together" [Quintavella & Heine, 2019, p. 686]. Figure 15.2 shows how AI-powered devices and robots that constitute smart homes systems would fulfil basic needs according to the different level of

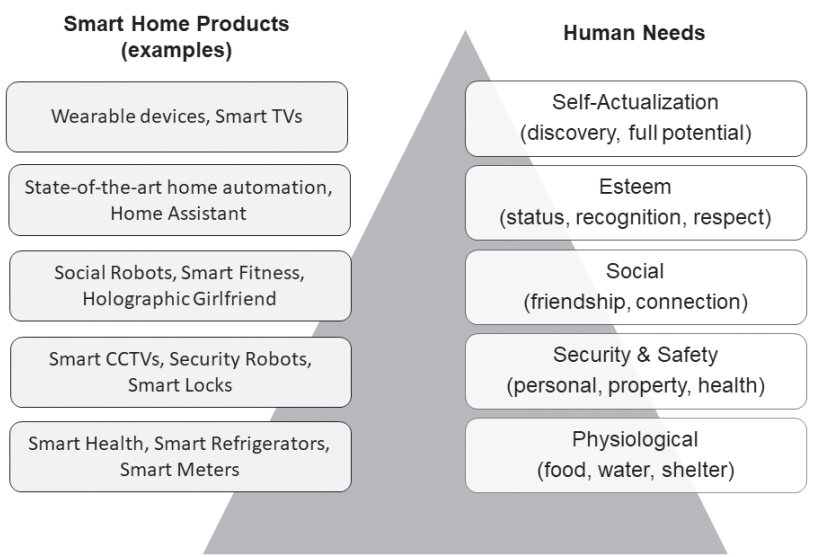



Figure 15.2. Smart homes devices *vis-à-vis* the Maslow Hierarchy of Needs.

*Source*: Adapted from Ovum 2015.

priorities that Maslow's hierarchy of needs prescribes [Ovum, 2015]. Indeed, the juxtaposition of smart devices with Maslow's hierarchy of needs illustrates that smart homes are not only objects of opulence; they can provide us with the necessary tools for human survival [Whatley, 2016].

Now, cyber defenders may ask how smart homes and their connection with human rights and people's basic needs are relevant to national security. Political scientists have argued that "national security and human rights are inextricably bound together" and that "… human rights, good governance … are themselves vital weapons in combating …" threats such as terrorism [Yasenchak *et al*., 2006, pp. 18 & 23]. Additionally, national security could exist only if the government has the political legitimacy to perform administrative functions such as law enforcement and military defence. People's trust is an essential indicator of political legitimacy. More importantly, the government could earn trust only through "concrete demonstrations of good intent" by fulfilling the basic needs that the Maslow hierarchy outlines [NIS Foundation, 2014]. During times of crisis or conflict, people's trust "empowers the government to act decisively" [Ho, 2018].

When threat actors compromise AI-powered devices and robots in smart homes, they could cause the problem of basic needs being unmet. Cyberattacks on smart homes that are persistent and harm-causing could shatter the occupants' or people's faith in AI and robots. When this occurs, people may demand that the government and its regulatory agencies do more to protect them from rogue AI and robots should they deem cybersecurity efforts by the industry deficient. The persistence of this situation may result in the perception that the government is not doing enough to meet people's basic needs. Psychologically, citizens may feel less confident of their government's capability to defend their human rights, which includes the fulfilment of their basic needs. The situation would become more challenging for the government if threat actors were to utilise adversarial AI to commandeer AI-powered devices and robots in smart homes. Over time, this could erode the people's trust and in turn, undermine the political legitimacy of the government.

## 15.3 Defence Approaches for Smart Homes

With the concept of home security evolving in the 4IR era, a home would soon require defending in both the physical space and cyberspace. It is thus unrealistic to expect smart homes to be impenetrable and to develop

defence measures that are only threat actor-specific. Instead, cyber defenders should develop deterrence approaches that aim to disrupt a spectrum of attacks and make threat actors give up on their attempts [Andress & Winterfeld, 2014]. This section will discuss two deterrence approaches that could defend smart homes from various threat actors.

### 15.3.1 *Approach #1: Deterrence by disruption*

The first approach is "deterrence by disruption," which comprises measures that would make it more challenging for threat actors to hack into smart homes. This approach would be useful for countries that find it difficult to attribute attacks, and to punish threat actors for their hostile actions due to operational, political, and geostrategic reasons. Countries may also choose not to retaliate as punitive measures could lead to conflict escalation with unbearable implications. Such constraints to "deterrence by punishment" are real, especially for countries that consider themselves as small states.

Two forms of active measures could serve "deterrence by disruption" by actively monitoring and hunting for possible hostile activities in the cyber environment of smart homes. The first form of active measure is the use of defensive AI to analyse voluminous data, detect anomalies, and respond to new and persistent cyber-attacks launched by adversarial AI (see Figure 15.3). Defensive AI could serve as an ally in counterattacks that pits protective machines against malicious machines and mirrors the human–machine partnership between adversarial AI and threat actors [Dixon & Eagan, 2019]. This partnership would be crucial as traditional cybersecurity measures are reactive and may not keep up when threat actors "change course and create new methods" [TCS, n.d., p. 79]. Defensive AI could surpass humans in terms of speed and scale of data analysis, especially when they are built with algorithms with advanced machine learning capabilities [Banham, 2018]. These capabilities would enable defensive AI to learn to "identify patterns that indicate malicious programmes from scratch" and proactively "take certain predefined steps in the event of an attack" [Bharadwaj, 2019a, pp. 2 and 4]. The use of defensive AI to defend smart homes would be comparable to the deployment of military robots and AI-driven battlefield systems in the era of machine-driven warfare.

The second form of active measures is the use of honeypots to entrap threat actors who are probing the cyber environment of smart homes for

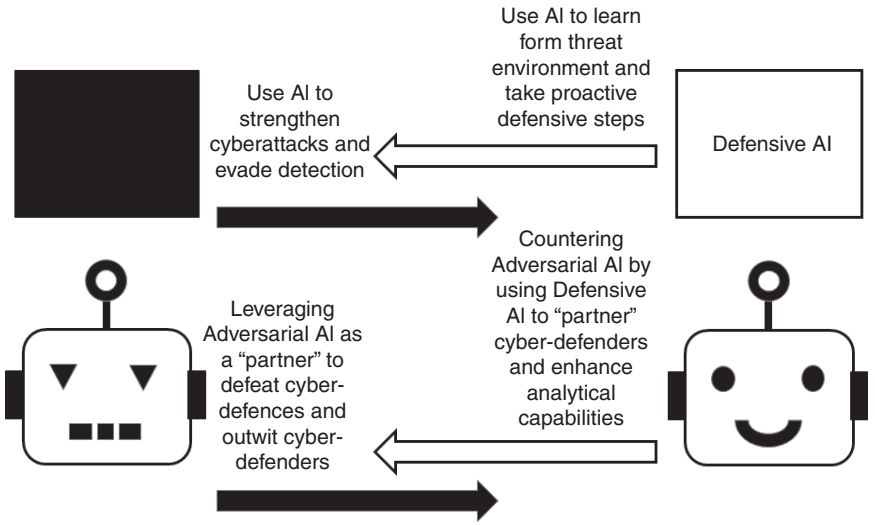



Figure 15.3. Illustration of Defensive AI countering Adversarial AI.

*Source*: Adapted from Rao 2019.

vulnerabilities. Honeypots are cyber tools that act as decoys to attract hostile activities [Andress & Winterfeld, 2014]. When a honeypot is hacked, cyber defenders could then analyse the behaviours that constitute the hostile activities and develop intelligence to take pre-emptive actions against the threat actors. The use of honeypots to defend smart homes would be comparable to the use of human agents (HUMINT) in traditional intelligence gathering for national security.

By identifying the vulnerabilities that the threat actors are targeting, cyber defenders could take more specific cybersecurity measures [Piggin & Buffey, 2016]. Currently, honeypots focus mainly on traditional computer systems, but versions for networked robotic systems are being developed [Irvene *et al*., 2017]. For example, researchers at the Georgia Institute of Technology have developed the HoneyBot as a honeypot for smart factories [Locklear, 2018]. What the HoneyBot does is attract and detect hostile activities that attempt to hack and seize control of robots and IoT devices that perform tasks in a high-tech factory. Additionally, cybersecurity companies such as Kaspersky have demonstrated that honeypots are useful in detecting attacks on smart devices such as routers and security cameras [Bayern, 2019].

### 15.3.2 *Approach #2: Deterrence by denial*

The second approach is "deterrence by denial," which comprises measures to respond to cyberattacks by persistent threat actors. These measures seek to persuade threat actors that "the potential benefit they obtain from the damage inflicted or the intelligence they collect will be less than the effort and resources they need to execute the attack" [Tolga, 2018, p. 7]. This approach is useful if countries are unable to stop cyberattacks. Since the cyber environment of smart homes by nature could never be impenetrable, this approach serves as a second line of defence should "deterrence by disruption" fails. Fundamentally, this approach entails "building resilient cyber defence systems that include all hardware, software, policy and human factors" [Tolga, 2018, p. 16].

Cyber resilience often relates to the operational and technical aspects—hardware, software, and policy—of protecting and recovering from cyberattacks. These factors constitute the "hard" aspects of cyberspace. Concerning smart homes, cyber defenders should dedicate more attention to the human factor of resilience due to the following reasons. Firstly, the proximity between the occupants and compromised AI-powered devices and robots in smart homes could increase the risk of physical harm. Secondly, the cyber-attacks affecting smart homes would be too close for comfort. People's trust and confidence in their government's ability to defend them may decline. These psychological and social effects could weaken the government's political legitimacy. Thirdly, research suggests that people "are more likely to respond to the effects of a cyber-attack rather than the attack itself" [Bada & Nurse, 2020, p. 2]. People may direct their fears and frustration more towards their government and each other instead of the threat actors. Therefore, attention to the human factor of resilience is crucial to preserving social cohesion and trust between the government and people, which constitute the "soft" aspects of cyberspace. These "soft" aspects must be robust for a country to effectively harness its source of strength and will—the people—when facing crisis and conflict.

"True resilience" is perhaps the most appropriate way of defining the human factor. Research suggests that people should imbibe three characteristics to achieve "true resilience." These characteristics are the (1) capacity to accept and face down reality; (2) ability to find meaning in some aspects of life; and (3) ability to improvise [Coutu, 2002]. Additionally, "true resilience" requires "understanding just

how interconnected and interdependent the different segments of the organisation are, as well as the third parties they rely on" [Ferbrache, 2018, p. 1]. This organisational characteristic could apply to people with their society and particularly in understanding how they depend on each other in the course of their daily life and work.

Concerning smart homes, the characteristics of "true resilience" could manifest in three ways (see Figure 15.4). Firstly, people must accept that it is not a matter of if, but when a successful cyberattack could affect their smart homes. In highly digitalised societies, cyberattacks could become more frequent than kinetic terrorist attacks. Secondly, people should imbibe the narrative of "crisis as an opportunity" in their national psyche. Cyberattacks could help people to understand where their vulnerabilities lie and decide what they should do to come out stronger. Thirdly, people

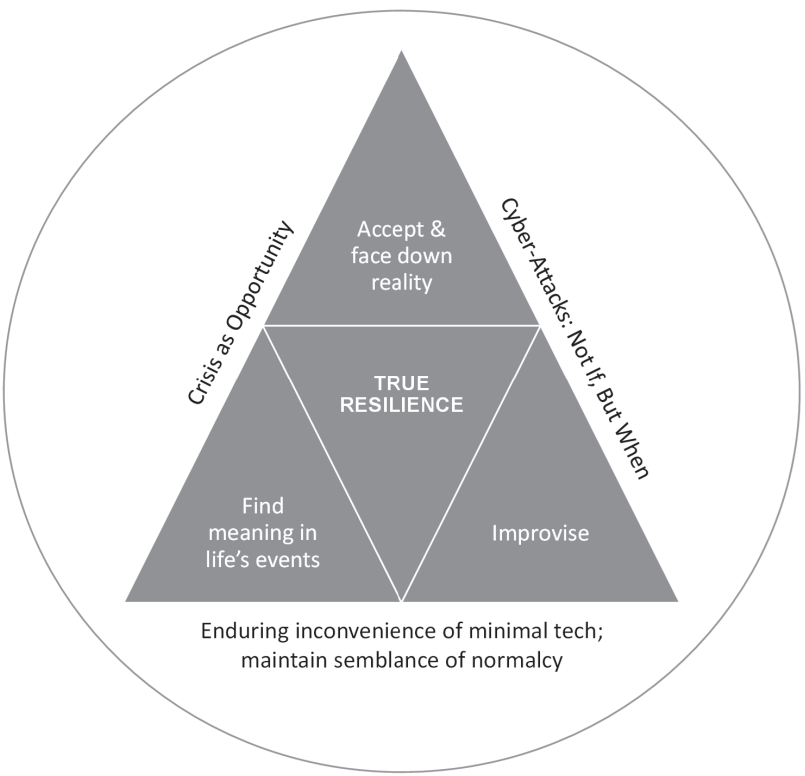



Figure 15.4. Applying the "True Resilience" model to defend against cyber threats that affect AI-powered devices and robots in smart homes.

should devise means to endure the inconvenience of life with minimal technology. Improvisation is crucial to maintaining a semblance of normalcy in life and work when threat actors compromise their smart homes [Faizal, 2020]. At the national level, incident response training and education that focuses on smart homes could enhance the people's preparedness by sharpening these three characteristics. Cyberattacks against smart homes would be less rewarding to threat actors if people who are the occupants could mitigate the effects and emerge stronger.

## 15.4 Conclusion

This chapter discussed that one space and context in the future that most people are likely to interact with AI and robots is their residence—smart homes. Research suggests numerous plausible ways that threat actors could wreak havoc in smart homes by compromising and turning AI-powered devices and robots against the occupants. As smart home technologies become increasingly available, it is necessary to understand the associated cyber threats and how to deter the threat actors [Barker, 2019]. This is especially so considering how smart homes could evolve into a national security issue given the far-reaching implications when compromised.

For smart homes that AI and robots operate, any minimal human oversight could result in occupants becoming belatedly aware of the security breaches (i.e., only when the threats materialise). Threat actors could exploit both the proximity between these machines and the occupants, and the environment of trust to cause harm. Motivations that could explain the act of hacking smart homes include curiosity, personal gain, terrorism, and geopolitical agendas. Geopolitical agendas pose the highest threat level, given the greater sophistication of cyberattacks and the potential impact on the targeted country's people. At the strategic level, people are the social centre of gravity of the targeted country as they are the source of national power—the nation's strength and will to fight. Additionally, cyberattacks on smart homes could undermine national security by depriving the people in the targeted country of their human rights and basic needs. There could be implications for the political legitimacy of the targeted country's government.

Cyber defenders should develop deterrence approaches that aim to disrupt a spectrum of attacks and make threat actors give up on their attempts. The first approach is "deterrence by disruption," which

comprises measures that could make it more challenging for threat actors to hack smart homes. The approach includes the use of active measures such as defensive AI and honeypots. The second approach is "deterrence by denial," which comprises measures to respond to cyberattacks by persistent threat actors. This approach primarily entails building up cyber resilience, and when concerning smart homes, the human factor of resilience is paramount. Together, both deterrence approaches constitute a "Defence-In-Depth" risk management philosophy for smart homes [Defence Online, 2019].

## 15.5 Acknowledgement

The views expressed in this chapter are the author's only and do not represent the official position or view of the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University (NTU), Singapore.

## 15.6 References

Altaher, Nada., Hauser, Jennifer., & Kottasova, Ivana. (2019). Yemen's Houthi Rebels Claim a 'Large-scale' Drone Attack on Saudi Oil Facilities. *CNN World*. https://edition.cnn.com/2019/09/14/middleeast/yemen-houthi-rebels-drone-attacks-saudi-aramco-intl/index.html

Andress, Jason & Winterfeld, Steve. (2013). *Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners (Second Edition)*. Syngress.

Asimov, I. & McQuarrie, R. (1990). *Robot Visions*. Penguin Books.

Bada, Maria & Nurse, Jason R. C. (2020). The Social and Psychological Impact of Cyber-attacks. *Emerging Cyber Threats and Cognitive Vulnerabilities*. (Academic Press) pp. 73–92.

Banham, Russ. (2018). Using AI and Machine Learning to Anticipate Cyber Threats. *Perspectives, Dell Technologies*. https://www.delltechnologies.com/en-us/perspectives/using-ai-and-machine-learning-to-anticipate-cyber-threats/

Barker, Sara. (2019). Five Ways Attackers can Create Havoc in Smart Homes. *Security Brief*. https://securitybrief.asia/story/five-ways-attackers-can-create-havoc-in-smart-homes

Bayern, Macy. (2019). Kaspersky Honeypots Find 105 Million Attacks on IoT Devices in the First Half of 2019. *TechRepublic*. https://www.techrepublic.com/article/kaspersky-honeypots-find-105-million-attacks-on-iot-devices-in-first-half-of-2019/

Bharadwaj, Raghav. (2019a). Artificial Intelligence in Cybersecurity—Current Use—Cases and Capabilities. *Emerj Weekly*. https://emerj.com/ai-sector-overviews/artificial-intelligence-cybersecurity/

Bharadwaj, Raghav. (2019b). Artificial Intelligence in Home Robots—Current and Future Use-Cases. *Emerj Weekly*. https://emerj.com/ai-sector-overviews/artificial-intelligence-home-robots-current-future-use-cases/

Brundage, Miles *et al*. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation. *Future of Humanity Institute, Oxford*. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Bugeja, Joseph, Jacobsson, Andreas, & Davidsson, Paul. (2017). An Analysis of Malicious Threat Agents for the Smart Connected Home. *The First International Workshop on Pervasive Smart Living Spaces 2017*. http://muep.mau.se/bitstream/handle/2043/22578/07917623.pdf;jsessionid=240EF78F42FB170A640E02A1156269FA?sequence=4

Bursztein, Elie. (2017). Inside the Infamous Mirai IoT Botnet: A Retrospective Analysis. *The Cloudflare Blog*. https://blog.cloudflare.com/inside-mirai-the-infamous-iot-botnet-a-retrospective-analysis/

Chen, Brian X. (2020). The Tech That Will Invade Our Lives in 2020. *The New York Times*. https://www.nytimes.com/2020/01/01/technology/personaltech/tech-trends-2020.html

Coutu, Diane. (2002). How Resilience Works. *Harvard Business Review*. https://hbr.org/2002/05/how-resilience-works

Defence Online. (2019). What is Defence in Depth? *Defence Contracts Online*. https://www.contracts.mod.uk/do-features-and-articles/what-is-defence-in-depth/

Dermer, James B. (2013). Cyber Warfare: New Character with Strategic Results. *United States Army War College*. https://apps.dtic.mil/dtic/tr/fulltext/u2/a589312.pdf

Dixon, William & Eagan, Nicole. (2019). 3 ways AI will change the nature of cyber-attacks. *World Economic Forum*. https://www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/

Epstein, Adam. (2016). Mr. Robot Played to our Worst Technology Fears with a Mini Horror Movie About a Hacked Smart Home. *Quartz*. https://qz.com/733269/mr-robot-played-to-our-worst-technology-fears-with-a-mini-horror-movie-about-a-hacked-smart-home/

Faizal, Muhammad. (2020). Singapore Well-placed to Tackle Security Threats in 2020, but the Dark Side Looms. *Today Online*. https://www.todayonline.com/commentary/four-major-security-trends-will-impact-singapore-2020

Ferbrache, David. (2018). Could your Business Survive a Cyber-attack? *KPMG LLP*. https://home.kpmg/uk/en/home/insights/2018/03/could-your-business-survive-a-cyber-attack.html

Fortune, Business Insights. (2019). Home Automation Market to Reach US$114 Billion by 2025; Advent of IoT Unlocks Lucrative Prospects, says Fortune Business Insights. *Press Release, Fortune Business Insights*. https://www.fortunebusinessinsights.com/press-release/home-automation-market-9047

Georgia Tech. (2018). Robot Designed to Defend Factories Against Cyber threats. *Georgia Institute of Technology*. https://rh.gatech.edu/news/604462/robot-designed-defend-factories-against-cyberthreats

Gray, Colin. (2013). *Perspectives on Strategy*. Oxford University Press.

Guo, Xiao, Shen, Zhenjiang, Zhang, Yajing, & Wu, Teng. (2019). Review on the Application of Artificial Intelligence in Smart Homes. *Smart Cities*, 2019, *2*(3), pp. 402–420. https://www.mdpi.com/2624-6511/2/3/25/htm

Hawking, Stephen, Russell, Stuart, Tegmark, Max, & Wilczek, Frank. (2014). Stephen Hawking: Transcendence Looks at the Implications of Artificial Intelligence—But Are We Taking AI Seriously Enough? *Independent*. https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html

Heinl, Caitriona. (2016). The Potential Military Impact of Emerging Technologies in the Asia-Pacific Region: A Focus on Cyber Capabilities. In Bitzinger, Richard A. (Ed.), *Emerging Critical Technologies and Security in Asia-Pacific*. (Palgrave Macmillan) pp. 123–137.

Ho, Peter. (2018). Is the Balance of Trust Shifting from Political to Social. *The Straits Times*. https://www.straitstimes.com/opinion/is-balance-of-trust-shifting-from-political-to-social

Irvene, Celine, Formby, David, Litchfield, Samuel, & Beyah. (2017). HoneyBot: A Honeypot for Robotic Systems. *Proceedings of the IEEE*. https://www.researchgate.net/publication/320048448_HoneyBot_A_Honeypot_for_Robotic_Systems

Kharkar, Anant, Anderson, Hyrum S., Filar, Bobby, & Roth, Phil. (2017). Evading Machine Learning Malware Detection. *Black Hat USA*. https://www.blackhat.com/docs/us-17/thursday/us-17-Anderson-Bot-Vs-Bot-Evading-Machine-Learning-Malware-Detection-wp.pdf

Klein, Alyson. (2019). How to Spot a Teenage Cyber-Hacker. *Education Week*. http://blogs.edweek.org/edweek/DigitalEducation/2019/09/teenage-hack-delinquency-cybersecurity.html

Kobie, Nicole. (2018). To Cripple AI, Hackers are Turning Data Against Itself. *WIRED*. https://www.wired.co.uk/article/artificial-intelligence-hacking-machine-learning-adversarial

Launius, D. Roger & McCurdy, E. Howard. (2008). *Robots in Space: Technology, Evolution and Interplanetary Travel*. The John Hopkins University Press.

Lin, Patrick. (2017). Why Cyberattacks could be War Crimes. *World Economic Forum*. https://www.weforum.org/agenda/2017/07/why-cyberattacks-could-be-war-crimes/

Linder, Courtney. (2019). So Maybe These Hackable Hotel Robots Were Not the Best Idea. *Popular Mechanics*. https://www.popularmechanics.com/technology/robots/a29590119/hotel-robots-spying/

Lindsay, Grey, Woods, Beau, & Corman, Joshua. (2016). Smart Homes and the Internet of Things. *Atlantic Council Issue Brief*. https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/smart-homes-and-the-internet-of-things/

Locklear, Mallory. (2018). HoneyBot Lures Hackers to Protect their Fellow Robots. *Engadget*. https://www.engadget.com/2018/03/29/honeybot-lures-hackers-protect-fellow-robots/

Manku, Tajinder. (2018). AI and Smart Home Automation for Connected Living. *The Fast Mode*. https://www.thefastmode.com/expert-opinion/12492-ai-and-smart-home-automation-for-connected-living

Newman, Lily Hay. (2019). The Worst Hacks of the Decade. *WIRED*. https://www.wired.com/story/worst-hacks-of-the-decade/

NIS Foundation. (2014). Think Piece. *NIS Foundation*. https://nis-foundation.org/wp-content/uploads/2018/05/NIS_Think_Piece_October_2014.pdf

NSI. (2019). Commodification of Cyber Capabilities: A Grand Cyber Arms Bazaar. *NSI*. https://nsiteam.com/social/wp-content/uploads/2019/11/191119-AEP_Commodification-of-Cyber-Capabilities-Paper.pdf

Ovum. (2015). Smart 2025: The Future of the Connected Home and Community. *Ovum*. https://www.broadband4europe.com/smart-2025-future-connected-home-community/

Piggin, Richard & Buffey, Ian. (2016). Active Defence Using an Operational Technology Honeypot. *Atkins*. https://www.atkinsglobal.com/~/media/Files/A/Atkins-Corporate/uk-and-europe/services-documents/cyber/Active%20defence%20with%20an%20OT%20honeypot.pdf

Porter, Christopher. (2018). In Cyber Warfare, the Front Line is Everywhere the U.S. Government Isn't. *Lawfare*. https://www.lawfareblog.com/cyber-warfare-front-line-everywhere-us-government-isnt

Quintavella, Alberto & Heine, Klaus. (2019). Priorities and Human Rights. *The International Journal of Human Rights*, *23*(4), pp. 679–697. https://doi.org/10.1080/13642987.2018.1562917

Rao, Joysula. (2019). Detection and Mitigation of Adversarial Attacks and Anomalies: Using AI for Security and Securing AI. *Robust Machine Learning Algorithms and Systems for Detection and Mitigation of Adversarial Attacks and Anomalies: Proceedings of a Workshop*, The National Academic Press, p. 14. https://www.nap.edu/read/25534/chapter/5#14

Sholihyn, Ilyas. (2019). Roomba Gets Modded to Swear, Curse, and Question Existence When it Hits Things. *Asia One*. https://www.asiaone.com/digital/roomba-gets-modded-swear-curse-and-question-existence-when-it-hits-things

TCS. (n.d.). A Machine First Approach to Digital Transformation. *Perspectives, Tata Consultancy Services*, 12. https://sites.tcs.com/bts/p12-perspectives-machine-first/

Tolga, Ihsan Burak. (2018). Principles of Cyber Deterrence and the Challenges in Developing a Credible Cyber Deterrence Posture. *NATO Cooperative Cyber Defence Centre of Excellence*. https://ccdcoe.org/uploads/2018/10/Challenges_in_Developing_Credible_Cyber_Deterrence_Posture_in_Cyberspace-1.pdf

Toon, John. (2019). Researchers have Chosen Which Smart Devices are at Risk of being Hacked. Here are the Results. *World Economic Forum*. https://www.weforum.org/agenda/2019/09/smart-devices-are-a-gateway-for-security-risks-heres-how/

Trend Micro. (2019). The New Norm: Trend Micro Security Predictions 2020. *Trend Micro Research*. https://www.trendmicro.com/vinfo/us/security/research-and-analysis/predictions/2020

UN. (2019). Keynote Speech: Human Rights in the Digital Age—Can they Make a Difference? *United Nations Human Rights Office of the High Commission*. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25158&LangID=E

Veer, Renske van der. (2019). Terrorism in the Age of Technology. *Netherlands Institute of International Affairs*. https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terrorism-in-the-age-of-technology/

Vincent, James. (2017). Home robots can be easily hacked to spy on and attack owners, say, researchers. *The Verge*. https://www.theverge.com/2017/8/22/16183514/hack-home-robot-surveillance-ioactive

Whatley, Jason. (2016). The 'Hierarchy' of Needs and the Connected Home. *Centralite* Blog. https://centralite.com/node/75

Wong, Pei Ting. (2019). The Big Read: Dangers Lurk in Youth's Chase for Social Media 'Likes'. *Channel News Asia*. https://www.channelnewsasia.com/news/singapore/instagram-tiktok-social-media-danger-gen-z-youth-the-big-read-11760046

Yacine, Ait Kaci. (2015). *Universal Declaration of Human Rights/Illustrations by YAK*. United Nations. https://www.un.org/en/udhrbook/pdf/udhr_booklet_en_web.pdf

Yasenchak, Megan A., Giglio, Jennifer, & Paxson, Margaret. (2006). National Security and Human Rights. *Conference Proceedings, Woodrow Wilson International Center for Scholars*. https://www.wilsoncenter.org/sites/default/files/KI_G8.pdf

Yusof, Zaihan Muhammad. (2016). When Tweens Become Hackers: Children as Young as Ten Getting Involved in Cybercrime. *The Straits Times*. https://www.straitstimes.com/singapore/when-tweens-become-hackers-children-as-young-as-10-getting-involved-in-cybercrime

Zachary, Fryer-Briggs. (2019). Coming Soon to a Battlefield: Robots That Can Kill. *The Atlantic*. https://www.theatlantic.com/technology/archive/2019/09/killer-robots-and-new-era-machine-driven-warfare/597130/

Zhang, Baobao & Dafoe, Allan. (2019). Artificial Intelligence: American Attitudes and Trends. *Center for the Governance of AI, Future of Humanity Institute, University of Oxford*. https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/general-attitudes-toward-ai.html

Ziska, Fields (Ed.). (2018). *Handbook of Research on Information and Cyber Security in the Fourth Industrial Revolution*. IGI Global.

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0016

# Chapter 16

# Understanding and Mitigating the Risk of Hackercide

Karthigan Subramaniam

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Karthigan_Subramaniam@mha.gov.sg*

## 16.1 Introduction

Today, smart devices are in high demand. The Internet of Things (IoT)[a] is one of the most popular catchphrases in the technology industry and with cities racing to digitise societies with fifth-generation mobile networks (5G), IoT devices will only become more common. There are approximately more than 21 billion IoT devices worldwide in 2020 and these numbers are expected to double by 2025 [World Economic Forum, 2020]. As more physical objects are becoming smart and connected with the rapid proliferation of IoT, threats are becoming more sinister with widespread repercussions for governments, businesses, and everyday users. Under these circumstances, the World Economic Forum [2020]

[a]IoT refers to internet-connected physical devices, objects, vehicles, buildings, and other items, which are able to collect and exchange data using embedded sensors that a user can monitor and/or control remotely [Staalduinen & Joshi, 2019]. Some of these examples would include driverless cars and smart refrigerators.

highlighted that this emerging technology could potentially make society more vulnerable to cyberattacks.

Most of the well-known hacking incidents so far have been confined to the digital world and have involved little more than stealing of data. However, the fact that these systems are connected to the cyberspace implies that they can be potentially hacked and manipulated for purposes other than what they were originally intended for [Wang *et al*., 2015]. According to Ivezic [2018], these kinds of cyberattacks are specifically known as cyber-kinetic attacks, and refer to either a direct or indirect exploitation of vulnerable interconnected systems to cause immense real-world ramifications such as physical damage, injury, or death. Such attacks have already occurred with physical damage inflicted on a variety of critical infrastructure such as nuclear power plants, water facilities, oil pipelines, factories, hospitals, transit systems, and apartment structures. More recently, hackers have also begun targeting smart devices within homes.

Table 16.1 presents a small sample of cyber-kinetic attacks that have occurred around the world. While there have been instances of damage and injury as shown in Table 16.1, the IoT industry has avoided cyber-kinetic attacks resulting in death, mainly due to a number of factors such as the inexperience of the attackers, the motive of the attackers (e.g., financial gain and thrill), and the quick work of responders. Although there have not been any reported cases of cyber-kinetic attacks with the intent to commit murder thus far, no nation is immune to this phenomenon. Similar to how crimes such as fraud and harassment have found new forms in cyberspace, it is only a matter of time before technology-facilitated murder via cyber-kinetic attacks become a reality. Hence, it is crucial that readers grasp the clear and present danger posed by the hacking of IoT devices to commit murder and take appropriate action to reduce this threat.

Against this backdrop, this chapter will focus on murder as a result of the hacking of IoT. For the purpose of this chapter, the act of committing murder via the hacking of IoT will be referred to as hackercide. It should also be noted at the outset that discussing the hacking of Internet of Military Things (IoMT) for cyberwarfare such as Unmanned Aerial Vehicles (UAV) is beyond the scope of this chapter. Instead, this chapter calls attention to the phenomenon of hackercide as a form of cybercrime, and attempts to provide a realistic assessment as to why such an incident has not yet occurred. Three measures that can help to further mitigate this

Table 16.1. Examples of cyber-kinetic attacks.

| Year | Incident |
|---|---|
| **2000** | An Australian wastewater engineer remotely accessed parts of the Maroochy Shire wastewater equipment after he was terminated and *released hundreds of thousands of litres of raw sewage into local parks and rivers throughout the town* |
| **2006** | A pair of LA traffic engineers attacked the Los Angeles traffic control system, *snarling traffic for days* |
| **2008** | A Polish teenager took control of a city's tram system and derailed four trams causing *multiple rider injuries* |
| **2010** | The Stuxnet worm destroyed uranium enrichment centrifuges in an Iranian nuclear power plant, *affecting the well-being of 2,500 homes located within a nine-kilometre radius from the facility* |
| **2015** | The BlackEnergy attack on the Ukrainian power grid *left 80,000 consumers without power* |
| **2016** | The Mirai botnet launched on service provider Dyn caused *major US websites including Twitter, the Guardian, Netflix, Reddit, and CNN, to be down for a day* |
| **2016** | Hackers infiltrated a water treatment plant's control system and *tampered with the levels of chemicals used to treat tap water* |
| **2016** | An attack on apartment buildings in Finland via smart thermostats *left residents without heat or water in the middle of Scandinavian winter for nearly a week* |
| **2016** | Iranian hackers breached the remote controls of the small Bowman Dam in Rye Brook, New York. The dam was offline for repair and therefore, could not be accessed remotely, but the implications are alarming as the hackers *might have been attempting to take control of an identically-named dam in Oregon, which is a dreadful 75-metres tall and 243-metres long* |
| **2017** | Multiple hospitals had to shut down critical equipment or postpone operations during the WannaCry ransomware attack as *hospitals and clinics were forced offline* |
| **2018** | A baby monitor used to keep an eye on babies was hacked and the *hacker threatened to kidnap the baby* |
| **2019** | After gaining access into their smart-home devices, a Wisconsin couple was terrorised by a hacker over a 24-hour period where he *turned up the heat and spoke to them through a camera* |
| **2019** | A hacker managed to access a Mississippi family's Ring security camera and *encouraged their eight-year-old daughter to destroy her room* |

*Source*: Compiled by author from public sources.

risk will also be proposed. While there are other ways that the internet facilitates murder (e.g., using the dark web to hire contract killers), those will not be covered in this chapter.

## 16.2 Hackercide

Alongside unprecedented advances in digital technologies, the 21st century also brings with it entirely new methods of committing murder. The concept of murder through the hacking of IoT is interesting, given that unlike traditional murder where the killer needs to be physically present to carry out the heinous act, a hacker can now commit murder remotely via IoT devices, as shown in Table 16.2.

In fact, an Israeli cyber expert had warned that the next 9/11 would be carried out with hackers infiltrating air traffic controls instead of suicide bombers [Staff, 2015]. While this may sound like a far-fetched idea to some, research has revealed the severity and likelihood of this possibility. In 2017, an official from the Department of Homeland Security (DHS) shared that his team was able to remotely hack into the controls of a Boeing 757 that the department had bought, and could access and gain control over some functions of the plane's systems [Steinbuch, 2017]. While the details of their hack and their research are classified, they do prove the real threat that can be posed by hacking interconnected devices. Besides the Boeing 757, security researchers have also found vulnerabilities in other interconnected devices such as smart cars and medical devices.

### 16.2.1 *Hacking of smart cars*

The expanding capabilities of smart cars is an exciting development in the automotive industry as an increasing number of vehicles are connected

Table 16.2. Difference between murder and hackercide.

| | **Definition** |
|---|---|
| **Murder** | When an individual kills another with the desire to harm that particular person. |
| **Hackercide** | When a hacker remotely uses an *interconnected device* to commit murder. Internet-connected devices could be any physical devices, objects, vehicles, buildings, and any other items. |

to IoT. However, researchers have demonstrated how easily smart cars can be hacked into to kill someone. In the 2013 DEF CON hacker conference, security researchers Charlie Miller and Chris Valasek showed how they hacked into internal computers on a Toyota Prius and a Ford Escape to take over the vehicle's steering and brake systems [Greenberg, 2013]. They were then able to jerk the vehicle's steering wheels, slam on the brakes, and even disable the brakes altogether, regardless of what the driver tried to do. Similarly, the aforementioned duo hacked a Jeep Cherokee in 2015 using its Uconnect system and gained control of critical functions such as control of the steering wheel but this time, without physical access to the vehicle itself [Rashid, 2018]. Given these demonstrations, it is entirely possible that a malicious hacker with the capabilities could hack into such smart cars to commit murder.

### 16.2.2 *Hacking of medical devices*

In a 2011 Black Hat conference, security researcher Jerome Radcliffe gave a presentation on how continuous glucose monitors (CGM) and glucose pumps can be infiltrated by hackers and used for lethal attacks against the user. He hypothesised that it was possible to hack the CGM device, tricking users into thinking that their blood sugar levels are higher or lower than it actually is, causing unwitting users to administer a greater or smaller dose of insulin into their bodies [Radcliffe, 2011]. Similarly, for an attack on the insulin pump, hackers can alter the amount of insulin entering the body and send the user into insulin shock, possibly leading to death [Radcliffe, 2011]. The following year, Barnaby Jack, a security researcher with McAfee, translated Radcliffe's ideas into reality by successfully hacking into an insulin pump and having it empty all of its content into a see-through mannequin from 90 metres away, displaying how vulnerable such interconnected devices truly are [Bates, 2012].

Unfortunately, these vulnerabilities are not limited to CGMs and insulin pumps. In 2017, researchers discovered over 8,600 vulnerabilities in pacemakers that hackers could exploit to potentially harm or kill heart patients [Khandelwal, 2017]. In fact, besides the hacking of insulin pumps described above, Jack was also able to hack a pacemaker remotely and instruct it to send out a deadly 830-volt jolt [Kirk, 2012]. To make matters worse, Jack shared that it was also possible to create a computer worm with the ability to infect multiple pacemakers by spreading through their system like a virus committing mass murder [Kirk, 2012]. More recently,

in 2018, Billy Rios and Jonathan Butts discovered similar vulnerabilities in pacemakers that can allow nefarious hackers to control them remotely and deliver deadly shocks at will or deny a life-saving shock when a patient requires one [Newman, 2018].

While no death has resulted from cyber-kinetic attacks as of yet, the threat of hacking into IoT devices to cause death looms on the horizon. Indeed, a threat assessment published by the Europol had predicted that the first murder via hacking of devices connected to the internet would occur by the end of 2014 [Peachy, 2014]. Fortunately, six years later, there has been no reported murders via hacking of interconnected devices.

### 16.2.3 *No reported attacks so far*

There could be several reasons to explain why there are no reports of hackercide thus far. First, such deaths, if occurred, could be falsely attributed to a technological glitch or failure of the device instead of the actions of a hacker. A fundamental concept in both cybersecurity and digital forensics is the fact that identifying the hackers involved in a cyberattack is at times a difficult task due to the sophisticated tactics that hackers employ [Wheeler & Larsen, 2003]. Hackers employ various techniques and tools at their disposal, such as anonymous servers and The Online Router (TOR), to cover their digital tracks and as a result, investigators may not suspect foul play [Wang *et al*., 2015]. Research indicates that the increasing number of new IoT devices are not supported by existing digital forensic tools and methods, making it difficult for investigators to extract data from them without the support of a forensic advisor with specialised knowledge in the area of IoT [Servida & Casey, 2019]. In a similar vein, it is entirely possible that eyewitnesses of a murder might unwittingly attribute the murder to an accident or even a technical glitch, and unknowingly steer the investigations in the wrong direction for an investigator who is unfamiliar with such technology.

A second and more probable reason, is that there are precautions in place to prevent such cases of hackercide from occurring. For example, there has been early detection of such flaws and calls for action by researchers and cybersecurity experts using conferences such as Black Hat USA [Kumar, 2017]. In 2017, the US Food and Drugs Administration ordered the recall of about half a million pacemakers amid concerns that hackers could run the batteries down or alter the patients' heartbeats, resulting in their death [Hern, 2017]. Similarly, Fiat Chrysler Automobiles

recalled 1.4 million vehicles due to the software defect mentioned above which allowed Charlie Miller and Chris Valasek to take remote control of the Jeep Cherokee and even turn off its engine [Goldman, 2015].

In light of potential security concerns, other organisations have also switched to manual updates to prevent such potential hacks from occurring. For example, in October 2018, Medtronic's pacemakers discontinued their internet-based software update system, switching to manual updates instead after researchers found vulnerabilities in the system that could allow a hacker to install malicious codes [Nichols, 2018]. In fact, former US Vice-President Dick Cheney revealed that his doctors had seemingly requested for manufacturers to disable his pacemaker's Wi-Fi more than a decade earlier in 2007, out of similar security concerns [Peachey, 2014].

## 16.3 Potential Mitigating Approaches to Counter the Threat Posed by Hackercide: Singapore as a Case Study

Despite knowing that such cyber-kinetic attacks have possibly not occurred, it is crucial to be proactive and take necessary measures to further mitigate the risk of hackercide. More broadly, the potential mitigating approaches covered in this chapter aim to reduce and, if possible, prevent the use of IoT as a tool not just for hackercide but for cybercrime as a whole.

### 16.3.1 *The need to address psychological vulnerabilities in human behaviour*

The first and most important step in countering the threat of hackercide requires addressing people's laid-back mindset towards the risk of hackercide. Significantly, humans have frequently been regarded as the 'weakest link' in cybersecurity. We often look for cognitive shortcuts known as heuristics to reduce the amount of effort the human brain expends during decision-making. Heuristics can, however, result in cognitive biases. By identifying fundamental psychological vulnerabilities in human behaviour, one can begin to understand why humans tend to brush off the risk of emerging threats such as hackercide and manage that risk effectively.

One such bias relevant to our discussion on hackercide would be the optimism bias[b] [Schwarcz, 2018]. When it comes to cybersecurity, many individuals tend to dismiss the current landscape as too complex to truly understand the risks and believe that these risks do not apply to them. Nevertheless, the need for individuals to understand the threat landscape and to mitigate the risk of hackercide has grown exponentially over the years with the increased adoption of IoT. Similar perceptions were also shared in the Singapore context. Results from a 2018 Public Awareness Survey of 1,105 respondents revealed that more than half of the respondents felt that they would not be targets of cyberattacks such as malware [Loh, 2019]. Most individuals tend to prioritise ease of connection over security, wrongly believing that hackers are only concerned with high-profile targets like political figures and major corporations. 'Why would they target us?' seems to be the general opinion, but not only is this perspective naïve, it is also outdated. Smart devices these days are so interconnected that any system with vulnerabilities can be hacked to commit hackercide. While no case of hackercide has been reported in Singapore as of yet, it would be a mistake to live with a false sense of security and to confuse our cybersecurity with the physical security that we enjoy in Singapore.

Another reason for the nonchalance towards hackercide is that people tend to discount the possibility of unprecedented risks. Indeed, many did not expect the COVID-19 pandemic either [Halliburton, 2020]. This is known as the availability bias where individuals predict the future by looking at accessible past events and thereby often fail to adequately calculate the likelihood or risk of an event [Schwarcz, 2018]. However, this does not mean that IoT users are delusional. Personal experiences play a crucial role in grasping the risk of such threats [Osberg & Shrauger, 1986] and for now, the threat of hackercide might seem distant and unrelatable for some. In other words, without experiencing it, the threat of hackercide can be hard to grasp.

Such cognitive biases that lead to a risk-tolerant attitude can be overcome by making an event more accessible to individuals, such as by exposing them to concrete instances of the event's occurrence. In 2019, Japan took the radical step of hacking its own citizens in an attempt to test the nation's vulnerability to cyberattacks and alert its citizens to the risk

[b]Optimism bias is a cognitive bias that causes someone to believe that they are at lesser risk of experiencing a negative event compared to others.

posed by their IoT devices [Griffiths, 2019]. In the local context, Singapore has recently explored the use of interactive films in its public education strategy, albeit in the context of drug abuse, with its interactive anti-drug short film titled "High" [Loh, 2020]. A similar approach of using interactive films can be adopted for hackercide where every decision made by the protagonist—or rather the viewer on his behalf—takes him down a different path which can either be safe or have life-threatening consequences. Such interactive films provide an opportunity for the viewer to virtually experience the consequences of their laxity and consequently act as an impetus for a change in behaviour when it comes to securing IoT devices to prevent hackercide.

### 16.3.2 *The need to raise cyber hygiene practices*

Similarly, there is a greater need for public awareness of cyber hygiene practices such that individuals and corporations can effectively protect themselves in cyberspace [Ministry of Home Affairs, 2016], lowering their chances of becoming hackercide victims. Be that as it may, Vishwanath *et al.* [2020] revealed that there has been no clear definition of cyber hygiene, and defined it as the practices that online consumers should follow in order to protect their personal information and devices from cyberattacks.

Woefully, cyber hygiene practices in Singapore seem to be lacking. A survey by Singapore's Infocomm Media Development Authority [2017] revealed that while 80% of internet users had previously installed anti-virus software and security updates on their home computers, only approximately 32% of smartphone users had installed anti-virus software on their smartphones. Among those who did not install an anti-virus software on their smartphone, most of them either did not think that it was necessary to install an anti-virus software on their smartphones or were unaware that such forms of protection were necessary or even available [Infocomm Media Development Authority, 2017]. Similar trends were observed globally as well. Results from a survey of 4,000 respondents, consisting of 2,000 Americans and 2,000 Canadians, reported that while half of them owned between one to five IoT devices (e.g., smart thermostats, smart fridges, etc.), only approximately a third of them are concerned about their home networks being hacked via such devices [ESET, 2019].

With this in mind, the public need to play their part in improving the security of their IoT devices by adopting better cyber hygiene

Table 16.3. Some key cyber hygiene practices to mitigate hackercide.

| | |
|---|---|
| **IoT device purchasing** | • Research on the security of IoT products before purchasing<br>• Decide whether the IoT product in question is necessary for its intended purpose<br>• Buy IoT devices that have passwords and that allow for the default password to be changed |
| **IoT device set-up and maintenance** | • Change default password on products, networks, and services<br>• Use strong passwords on products, networks, and services<br>• Ensure that your Wi-Fi is secured<br>• Enable firewall on IoT devices if possible<br>• Restrict who can connect to your IoT device<br>• Check your IoT device to ensure it has the latest operating system, software update, or patch |

practices (see Table 16.3). In Singapore, the Cybersecurity Agency (CSA) introduced the Cybersecurity Labelling Scheme in 2020 as part of their efforts to raise cyber hygiene levels by providing an indication of the level of security that is embedded in the IoT devices [Tham, 2020]. While no case of hackercide has been reported as of yet, it is a timely reminder that security starts with each individual user's responsible behaviour in the cyber world.

### 16.3.3 *Making security part of the design of IoT*

Fighting any form of cybercrime, including hackercide, will always be a collaborative effort between individuals as well as the IoT developers. Unfortunately, there may be a lack of knowledge and awareness about the risks posed by IoT devices, as these smaller companies that focus on manufacturing smart-home devices such as smart thermostats and smart refrigerators may not be able to afford large teams of cybersecurity experts to ensure the security of their devices. As pointed out by Blythe *et al.* [2020], manufacturers have little incentive to ensure that their devices are secure by design. With companies competing to get their products out into market cheaply and quickly, software engineers routinely fail to incorporate security into their designs [Blythe *et al*., 2020]. Considering that their competitor's security is just as negligent as theirs, it would be a competitive disadvantage to prioritise the security of their own IoT devices. Unsurprisingly then, 80% of IoT applications are not

tested for vulnerabilities, primarily due to rushed release schedules [Ponemon Institute, 2017].

On this issue, minister-in-charge of the Smart Nation Initiative in Singapore, Dr. Vivian Balakrishnan, emphasised that IoT developers need to ensure that security is part of their design process and not scramble to tackle such issues after it has rolled out [Tanoto, 2018]. Security professionals must also address the new cyber-kinetic risks that IoT create such as hackercide and can consider supporting responsible security research by employing white hackers[c] to test their system, especially when certain IoT devices have been proven to be capable of hackercide (e.g., medical devices and smart cars) by other security analysts. Such security research is sorely lacking in general—only 13% of companies making IoT products have a public disclosure policy that allows researchers to probe known vulnerabilities [IoT Security Foundation, 2020]. Given that IoT systems are vulnerable to cyber-kinetic attacks, traditional security protocols and testing processes must be rethought and revised to catch up with current and emergent technologies like 5G.

In Singapore, at the government level, the CSA works closely with the Smart Nations and Digital Government Office (SNDGO) and the Government Technology Agency (GovTech) to ensure security-by-design, whereby cybersecurity is incorporated from the onset [Tan & Yimin, 2018]. For example, in 2018, GovTech and CSA launched the Government Bug Bounty Programme to work with both local and international white-hat hackers to identify and address vulnerabilities in selected government websites and digital services [Kwang, 2018]. A year later, the Vulnerability Disclosure Programme was launched as a complementary programme to provide a channel for members of the public to report vulnerabilities they discover on government websites or applications [Baharudin, 2019]. Likewise, several initiatives have been launched around the world, including in the UK government, European Union, and US government, in an effort to make IoT devices more secure [Palmer, 2018], which would in turn reduce the likelihood of IoT devices being hacked.

---

[c] White hackers, or ethical hackers, are hackers who have permission from the owner of the system they intend to hack and at times, can be paid employees working for companies as security specialists.

### 16.3.4 *Proactive advancement of law enforcement training and legislation*

Hackercide has implications not just for individuals and IoT developers but also in the way law enforcement personnel address these cases. It would be vital for investigation officers to consider the possibility of such forms of murder occurring and more importantly, to proactively develop digital or even IoT forensic techniques that could identify and trace the hackers. In Singapore, the Singapore Police Force (SPF) Cybercrime Command develops the curriculum for SPF's specialised cybercrime investigation training modules [Ministry of Home Affairs, 2016]. SPF has also embarked on several new technological initiatives to improve its cybercrime investigation capabilities [Ministry of Home Affairs, 2016]. Nevertheless, the modus operandi of cybercrime will continue to evolve, and our investigation processes have to keep abreast of these changes.

In addition, a robust criminal justice framework must support investigation of such cybercrimes. Cyber laws need to be constantly revised to be relevant and effective in deterring new forms of cyber-kinetic attacks such as hackercide. The idea of regulating IoT devices is gaining traction in some countries, including Singapore, albeit in the context of data protection and privacy as explained in further detail below. Having said that, most countries still lack a formal IoT framework [Greengard, 2019]. Although laws have often taken time to catch up to technology, situations like these could be a literal matter of life and death. Therefore, criminal justice processes must be quick and efficient in dealing with such emerging forms of cybercrimes. The importance of revising legislation was also underscored by Dr. Vivian Balakrishnan, who noted the need to rethink current rules and regulations with regard to IoT and the need to change policies to cope with this digital imperative [Wong, 2019].

In early 2020, both California and Oregon introduced new legislation requiring "reasonable security features" to be added to IoT devices [Merken, 2019]. While it is a commendable effort in the right direction by these states, both legislations ultimately lack specific instructions for organisations to comply with the new requirements as neither law precisely defines the term "reasonable," leaving it open to interpretation [Merken, 2019]. Incidentally, the UK government announced its plans to introduce new mandatory requirements for IoT manufacturers in early

2020 as well but provided clear guidelines such as disallowing default passwords, implementing a vulnerability disclosure policy, and keeping the device software updated in an effort to secure IoT devices [Department for Digital, Culture, Media and Sports, 2019]. In Singapore, the Computer Misuse and Cybersecurity Act (CMCA) enables the Ministry of Home Affairs to take pre-emptive measures rather than just countermeasures to prevent cyberattacks on critical infrastructures and essential services. While the CMCA was amended and expanded upon in 2017, by criminalising acts that create significant risk of harm in Singapore as well as the act of obtaining hacking tools to commit cyber offences, it is inadequate in addressing the vulnerabilities of personal IoT devices such as smart homes and smart vehicles. With this in mind, the Infocomm Media Development Authority (2020) recently launched an IoT Cyber Security Guide to provide vendors and end-users with guidance on addressing the cybersecurity aspects of IoT systems, such as changing default passwords of IoT devices, and the use of firewalls and anti-malware software.

On the whole, it appears that the new regulations are coming from a good place. Governments are taking necessary action to protect consumers from IoT-related vulnerabilities that could potentially result in hackercide. Nevertheless, requiring basic security on individual IoT devices may be a misguided approach. Rather, more needs to be done by IoT vendors who have the ability to implement more sophisticated cybersecurity features in the design phase to address the vulnerabilities in IoT devices in a more holistic and comprehensive manner.

## 16.4 Conclusion

As we move towards the implementation of 5G technology, the number of IoT devices are bound to increase with 5G networks promising a new era in connectivity and seamless real-time engagement between IoT devices. For the first time, driverless cars and other autonomous objects will become a reality, opening up new possibilities for hackercide to occur. As explained in this chapter, with the whole world moving towards 5G and an increased adoption of IoT, there is an increased risk of cyber-kinetic attacks and, in particular, hackercide. While cases of hackercide have not been reported as of yet, we need to take proactive steps to mitigate the risk of hackercide from an individual, system, as well as a legal perspective.

The addition of digital defence as the sixth pillar of Total Defence in Singapore in early-2019 is timely, as it requires the collective actions of both individuals as well as systems to defend against emerging cyber threats [Ministry of Defence, 2019]. Emphasis, therefore, needs to turn towards the detection and prevention of such hacking incidents. This means that individuals should not be complacent about the likelihood of hackercide but instead, take necessary measures and precautions to protect themselves and their loved ones by being aware of these vulnerabilities. Manufacturers must factor in security when developing emerging technologies like the IoT in order to pre-empt and protect against malicious hackers. Relying on past security measures to prevent hackercide is not an option. With the advent of IoT developments and as people and devices become more connected, we may very soon find that our most basic and relied upon devices will quickly become our biggest liabilities from a security perspective. Hence, IoT developers need to get security right because the ramifications for getting it wrong will include the loss of life. While it is true that 5G technology and IoT will offer huge benefits, this emerging technology necessitates a greater focus on security.

## 16.5 Acknowledgement

The views expressed in this chapter are the author's only and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 16.6 References

Baharudin, H. (2019, November 19). Hackers to test 12 govt systems in bug bounty programme. *The Straits Times.* https://www.straitstimes.com/tech/hackers-to-test-12-govt-systems-in-bug-bounty-programme

Bates, C. (2012, April 10). Hackers 'can gain access to medical implants and endanger patients' lives'. *Daily Mail.* https://www.dailymail.co.uk/health/article-2127568/Hackers-gain-access-medical-implants-and-endanger-patients-lives.html

Blythe, J. M., Johnson, S. D., & Manning, M. (2020). What is security worth to consumers? Investigating willingness to pay for secure Internet of Things devices. *Crime Science*, *9*(1). http://dx.doi.org/10.1186/s40163-019-0110-3

Department for Digital, Culture, Media and Sports. (2019). *Consultation on the Government's regulatory proposals regarding consumer Internet of Things (IoT) security.* www.gov.uk/government/consultation-on-regulatory-proposals-on-consumer-iot-security/consultation-on-the-governments-regulatory-proposals-regarding-consumer-internet-of-things-iot-security

ESET (2019, October 8). ESET survey finds disconnect between consumer attitudes and actions toward connected home privacy. https://www.eset.com/sg/about/newsroom/press-releases1/products/eset-survey-finds-disconnect-between-consumer-attitudes-and-actions-toward-connected-home-privacy/

Goldman, D. (2015, July 24). Chrysler recalls 1.4 million hackable cars. *CNN.* https://money.cnn.com/2015/07/24/technology/chrysler-hack-recall/index.html

Greenberg, A. (2013, July 24). Hackers reveal nasty new car attacks—with me behind the wheel. *Forbes.* https://www.forbes.com/sites/andygreenberg/2013/07/24/hackers-reveal-nasty-new-car-attacks-with-me-behind-the-wheel-video/#9d17202228c7

Greengard, S. (2019). Deep insecurities: The internet of things shifts technology risk. *Commun. ACM*, *62*(5): 20–22. https://doi.org/10.1145/3317675

Griffiths, J. (2019, February 2). 'Internet of things' or 'vulnerability of everything'? Japan will hack its own citizens to find out. *CNN.* https://edition.cnn.com/2019/02/01/asia/japan-hacking-cybersecurity-iot-intl/index.html

Halliburton, B. C. (2020, March 19). COVID-19 is a black swan. *Forbes.* https://www.forbes.com/sites/forbesbooksauthors/2020/03/19/covid-19-is-a-black-swan/#6592e3967b4b

Hern, A. (2017, August 31). Hacking risk leads to recall of 500,000 pacemakers due to patient death fears. *The Guardian.* https://www.theguardian.com/technology/2017/aug/31/hacking-risk-recall-pacemakers-patient-death-fears-fda-firmware-update

Infocomm Media Development Authority. (2017). *Annual survey on infocomm usage in households and by individuals for 2017.* https://www.imda.gov.sg/-/media/Imda/Files/Industry-Development/Fact-and-Figures/Infocomm-Survey-Reports/HH2017-Survey.pdf

Infocomm Media Development Authority. (2020, March 13). IMDA Launches IoT Cyber Security Guide to Help Enterprise Users and Vendors Secure IoT systems. https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2020/IMDA-Launches-IoT-Cyber-Security-Guide-to-Help-Enterprise-Users-and-Vendors-Secure-IoT-Systems

IoT Security Foundation. (2020). *Consumer IoT: Understanding the Contemporary Use of Vulnerability Disclosure—2020 Progress Report.* https://www.iotsecurityfoundation.org/wp-content/uploads/2020/03/IoTSF-2020-Progress-Report-Consumer-IoT-and-Vulnerability-Disclosure.pdf

Ivezic, M. (2018, January 2). The tangible threat of cyber-kinetic attacks. *CSO.* https://www.csoonline.com/article/3245036/the-tangible-threat-of-cyber-kinetic-attacks.html

Khandelwal, S. (2017, June 5). Over 8,600 vulnerabilities found in pacemakers. *The Hacker News.* https://thehackernews.com/2017/06/pacemaker-vulnerability.html

Kirk, J. (2012, October 7). Pacemaker hack can deliver deadly 830-volt jolt. *ComputerWorld.* https://www.computerworld.com/article/2492453/pacemaker-hack-can-deliver-deadly-830-volt-jolt.html

Kumar, C. (2017). New Dangers In the New World: Cyber Attacks in the Healthcare Industry. *Intersect: The Stanford Journal of Science, Technology, and Society*, *10*(3), 1–15. https://ojs.stanford.edu/ojs/index.php/intersect/article/view/1087

Kwang. (2018, September 18). Government to launch bug bounty programme by this year: DPM Teo. *CNA.* https://www.channelnewsasia.com/news/singapore/cybersecurity-government-to-launch-bug-bounty-programme-dpm-teo-10730780

Loh, G. S. (2020, March 23). Singapore's first interactive film is a horror story about drug use. *CNA.* https://cnalifestyle.channelnewsasia.com/trending/singapore-first-interactive-film-drug-use-meth-royston-tan-12566584

Loh, V. (2019, February 23). The Big Read: As more cyber attacks loom, Singapore has a weak 'first line of defence'. *Today.* https://www.todayonline.com/big-read/big-read-more-cyber-attacks-loom-singapore-weak-first-line-defence

Merken, S. (2019, October 29). Connected Device Makers Face California, Oregon Security Laws. *Bloomberg Law.* https://news.bloomberglaw.com/privacy-and-data-security/connected-device-makers-face-california-oregon-security-laws

Ministry of Defence. (2019, February 15). *Fact Sheet: Digital Defence.* https://www.mindef.gov.sg/web/portal/mindef/news-and-events/latest-releases/article-detail/2019/February/15feb19_fs

Ministry of Home Affairs. (2016, July 20). *National Cybercrime Action Plan*. https://www.mha.gov.sg/docs/default-source/press-releases/ncap-document.pdf

Newman, L. H. (2018, September 8). A New Pacemaker Hack Puts Malware Directly on the Device. *Wired.* https://www.wired.com/story/pacemaker-hack-malware-black-hat/

Nichols, S. (2018, October 12). It's the real Heart Bleed: Medtronic locks out vulnerable pacemaker programmer kit. *The Register.* https://www.theregister.co.uk/2018/10/12/medtronic_pacemaker_programmer_security/

Osberg, T. M., & Shrauger, J. S. (1986). Self-prediction: Exploring the parameters of accuracy. *Journal of Personality and Social Psychology*, *51*(5), p. 1,044. https://doi.org/10.1037/0022-3514.51.5.1044

Palmer, D. (2018, March 15). IoT security warning: Cyber-attacks on medical devices could put patients at risk. *ZDNet.* https://www.zdnet.com/article/

iot-security-warning-cyber-attacks-on-medical-devices-could-put-patients-at-risk/

Peachey, P. (2014, October 5). Cyber crime: First online murder will happen by end of year, warns US firm. *The Independent.* https://www.independent.co.uk/life-style/gadgets-and-tech/news/first-online-murder-will-happen-by-end-of-year-warns-us-firm-9774955.html

Ponemon Institute. (2017). 2017 Study on Mobile and Internet of Things Application Security. https://www.ponemon.org/local/upload/file/Arxan%20Report%20Final%205.pdf

Radcliffe, J. (2011). Hacking media devices for fun and insulin: Breaking the human SCADA system. *Black Hat Technical Security Conference.* Las Vegas: Nevada.

Rashid, F. Y. (2018, May 25). Hacker History: The time Charlie and Chris hacked a Jeep Cherokee. *Decipher.* https://duo.com/decipher/hacker-history-time-charlie-chris-hacked-jeep-cherokee

Schwarcz, S. L. (2018). Regulating Complacency: Human Limitations and Legal Efficacy. *Notre Dame Law Review*, *93*, pp. 1,073–1,104. https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4773&context=ndlr

Servida, F., & Casey, E. (2019). IoT forensic challenges and opportunities for digital traces. *Digital Investigation*, *28*, pp. S22–S29.

Staalduinen, M. V., & Joshi, Y. (2019). The IoT security landscape: adoption and harmonisation of security solutions for the internet of things. https://www.csa.gov.sg/~/media/csa/documents/publications/iot_security_landscape/iot%20security%20landscape%20report.pdf

Staff, T. (2015, April 15). Next 9/11 will be caused by hackers, not suicide bombers, cyber expert warns. *The Times of Israel.* https://www.timesofisrael.com/hackers-will-cause-next-911-cyber-expert-warns/

Steinbuch, Y. (2017, November 14). Cybersecurity expert claims he was able to hack into a parked Boeing 757. *The New York Post*. https://nypost.com/2017/11/14/cybersecurity-expert-claims-he-was-able-to-hack-into-a-parked-boeing-757/

Tan, B., & Yimin, Z. (2018). *Technology and the City: Foundation for a Smart Nation.* Centre for LiveableCities. (Urban Systems Studies). https://www.clc.gov.sg/docs/default-source/urban-systems-studies/uss-technology-and-the-city.pdf

Tanoto, B. (2018, March 21). Government's duty to set open standards for Internet of Things Deployment: Vivian Balakrishnan. *CNA*. https://www.channelnewsasia.com/news/singapore/government-s-duty-to-set-open-standards-for-internet-of-things-10062284

Tham, I. (2020, March 4). Parliament: New cyber security label for smart devices. *The Straits Times.* https://www.straitstimes.com/politics/parliament-cyber-security-label-for-connected-devices-to-address-growing-concern

Vishwanath, A., Neo, L. S., Goh, P. Lee, S., Khader, M., Ong, G., & Chin, J. (2020). Cyber hygiene: The concept, its measure, and its initial tests.

*Decision Support System*, *128*, 113160. https://doi.org/10.1016/j.dss.2019.113160

Wang, P., Chin, J., & Khader, M. (2015). *Murder by Hacking: The Emerging Threat of Kinetic Cyber.* Home Team Behavioural Sciences Centre.

Wheeler, D. & Larsen, G. (2003). Techniques for Cyber Attack Attribution. Technical report, Institute for Defense Analysis.

Wong, L. (2019, March 27). Laws must keep pace with technical trends to safeguard citizen's privacy and rights, experts say. *The Straits Times.* https://www.straitstimes.com/tech/laws-must-keep-pace-with-tech-trends-to-safeguard-citizens-privacy-and-rights-experts-say

World Economic Forum. (2020). Global Risk Report 2020. World Economic Forum report. http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf

# Section G

# Special Chapter

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0017

# Chapter 17

# Sexting in Singapore: An Empirical Study

Joey Low[*,‡] and Majeed Khader[†,§]

[*]*Nanyang Technological University, Singapore*

[†]*Home Team Behavioural Sciences Centre, Ministry of Home Affairs, Singapore*

[‡]*joeylowjlz@gmail.com*

[§]*Khader_MAJEED@mha.gov.sg*

## 17.1 Introduction

The internet has changed the way people communicate. One characteristic of the internet—anonymity—allows users to say and do things in cyberspace that they will not usually do in other settings. This has been known as the online disinhibition effect [Suler, 2014]. Although '*benign disinhibition*' can lead to more open constructive conversations, '*toxic disinhibition'* can instead lead people to explore new risky activities that carry major consequences, and one such example is sexting.

In 2012, Amanda Todd, a 15-year-old Canadian, was persuaded by a stranger to send a nude sext of her flashing her breasts [Strohmaier *et al*., 2014]. After the photo was leaked on the internet by the stranger, Amanda was harassed at school and cyberbullied by her peers [Crimmins & Seigfried-Spellar, 2014]. Consequently, Amanda developed depression and anxiety, factors which eventually contributed to her suicide. In a similar case, Jessica Logan sent nude pictures of herself to her boyfriend,

and after their relationship fell apart, he circulated those pictures to other high school girls. Jessica was harassed and bullied, factors which again, contributed to her decision to take her life [Aldridge *et al*., 2013]. These case studies illustrate that sexting is a potentially harmful phenomenon and there is a need to enhance our understanding of it.

While the concept of sexting is not foreign to Singapore, there has been a lack of sufficient local studies in this area. Thus, this chapter aims to (1) introduce the concept of sexting; (2) investigate the sexting behaviours in the local Singapore setting; and (3) identify psychosocial risk and protective factors associated with sexting.

## 17.2 What is Sexting?

Sexting is a neologism made up of the words "sex" and "texting." Individuals would share sexually explicit photos and messages (i.e., sext) via mobile phone or the internet [Messer *et al*., 2013]. Specifically, these sexually explicit photos contain revealing or even nude photos that are self-produced by the sender [Doring, 2014]. There are many different definitions of sexting. For instance, Lee, Moak and Walker [2016] defined sexting as "electronically transmitting sexually explicit texts and materials to others" [p. 3], while Dake, Price, and Maziarx [2012] defined it as "sending, receiving or forwarding sexually explicit messages, or nude, partially nude, or sexually suggestive digital images of one's self or others via a cell phone, email, Internet or Social Networking Sites" [p. 2]. In this study, sexting is defined as the sharing of sexually suggestive messages and/or nude (i.e., naked breasts, genitals or bottom), or semi-nude (e.g., photos taken with only undergarments on or focused on clothed genitals) photos.

The act of sending and receiving sexually suggestive images and messages is facilitated by the introduction of mobile phones and the advert of the internet [Weisskirch & Delevi, 2011]. For instance, in the United States, the 2008 National Campaign to Prevent Teen and Unplanned Pregnancy survey found that 33% of 1,280 young adults (between the ages of 13 and 26) reported having sent or posted semi-nude or nude images of themselves [Crimmins & Seigfried-Spellar, 2014].

Despite sexting becoming a potential cause of concern, there have not been many studies that have looked into this phenomenon. Among the few studies done, it was found that sexting behaviours were more prevalent among adults as compared to adolescents [Klettke *et al*., 2014]. One of the

most common reasons people engage in sexting is to develop or maintain a romantic relationship [Korenis & Billick, 2014; Ouytsel *et al*., 2014]. Another important factor that leads to sexting behaviours has been peer pressure. Dake, Price, and Maziarz [2012] found that 23% of teens in the US engaged in sexting due to pressure from a friend, while 51% of teenage girls were pressured by boys to send sexually explicit messages.

### 17.2.1 *The importance of understanding the phenomenon of sexting*

Sexting is a risky behaviour that can lead to many negative consequences. One of which is the scenario where the sexts are disseminated beyond the intended recipient [Doring, 2014]. With the convenience of technology, once a sext has been sent, the sender has virtually no control over what the receiver intends to do with it. Receivers might choose to circulate the sext as revenge due to the end of a relationship, to gain peer approval, or even for fun [Ouytsel *et al*., 2014]. A study conducted by Strassberg, Rullo, and Mackaronis [2014] in the US found that among those who received a sext, 18.7% of them sent the pictures on to others. As in the cases of Amanda and Jessica, once these sexts were circulated, it led to other negative consequences such as harassment and cyber bullying. These could be detrimental to the victims, with psychological repercussions such as depression, anxiety and suicide.

The second negative consequence is that sexting is found to be associated with risky behaviours. Studies have found that individuals who engaged in sexting were also more likely to engage in other sexual behaviours [Korenis & Billick, 2014]. For example, Dake, Price, and Maziarz [2012] had found that sexting was highly associated with having multiple sexual partners, and engaging in oral, anal, or unprotected sex.

Lastly, there are legal consequences associated with sexting. In the US, adolescents who engaged in sexting could be charged with the production and distribution of child pornography [Aldridge *et al*., 2013]. Furthermore, perpetrators would be registered as sex offenders. For instance, Phillip Alpert, an 18-year-old from Florida, was convicted of a felony and registered as a sex offender for 25 years after he circulated a nude picture of his 16-year-old ex-girlfriend [Strohmaier *et al*., 2014].

In Singapore, those who engage in sexting could be charged using the Singapore Penal Code, Chapter 338: Undesirable Publications Act, which is "an Act to prevent the importation, distribution or reproduction

of undesirable publications and for purposes connected therewith" [Attorney-General's Chamber, 2015].

## 17.3 Understanding Sexting in Singapore

In 2015, a survey of 2,700 Singapore secondary school students found that 4.2% of upper secondary and 1.9% of lower secondary school students engaged in sexting [Tai, 2015]. There are also reported cases such as a local secondary school girl who was labelled a 'slut' after her boyfriend showed nude and provocative photos of her—which she sent to her boyfriend—to his friends [Teng, 2015].

Sexting is not foreign to Singapore but there has been a lack of sufficient local studies in this area. Given the potential negative consequences associated with sexting, this chapter aims to provide insights about sexting behaviours in the local Singapore setting, and identify psychosocial risk and protective factors associated with sexting. For the study, 205 Singaporean participants[a] (62 males and 143 females) were recruited—113 were undergraduates from a University in Singapore, and 92 participants were recruited via WhatsApp and Facebook.

## 17.4 Sexting Behaviours in the Local Singapore Setting

Information about the sexting behaviours in Singapore, such as the frequency of occurrence, motivation, and consequences of sexting, was collected. Out of the 205 participants, 113 participants (55.1%) reported having received at least one sext in their life. Specifically, 89 participants

[a]The age of the participants ranged from 18 to 36 years old ($M$ = 22.48, $SD$ = 2.79). The study was conducted using the Qualtrics survey platform. The participants were briefed about the nature of the study at the beginning of the study and reassured about the anonymity and confidentiality of the data collected. Informed consent was sought, and the participants proceeded on with the questionnaires. After completion of the questionnaires, the participants were debriefed. Additionally, due to the sensitive nature of the study, participants were also given helplines that they can call should they experience any psychological discomfort.

(43.4%) reported having received sexually suggestive text messages, 85 participants (41.5%) received text messages propositioning sexual activity, 62 participants (30.2%) received sexually suggestive photos self-produced by the sender, while 47 participants (22.9%) had received nude pictures self-produced by the sender.

Similar trends were found for sending behaviours where 72 participants (35.1%) have engaged in some form of sexting. Table 17.1 shows the descriptive statistics of sexting behaviours. Among which, 64 participants (31.2%) reported sending sexually suggestive text messages, 41 participants (20.0%) sent text messages propositioning sexual activity, 34 participants (16.6%) sent self-produced sexually suggestive photos, and 24 participants (11.7%) sent self-produced nude photos. An example of a sexually suggestive text message sent is "What would you like me to do to you in bed?" and an example of a text message propositioning sexual activity sent is "Give me a blowjob."

There were some significant gender differences found in the sexting behaviours. More males reported receiving sexually suggestive photos self-produced by the sender ($\chi^2$ (1) = 4.28, $p < .05$; 40.3% of males, 25.9% of females), sending sexually suggestive text messages ($\chi^2$ (1) = 6.29, $p < .05$; 43.5% of males, 25.9% of females), and sending text messages propositioning sexual activity ($\chi^2$ (1) = 8.35, $p < .01$; 32.3% of males, 14.7% of females). Additionally, it was found that relationship status was associated with sending sexts, where individuals who were in a relationship were more likely to have engaged in all four sexting behaviours. Individuals who were in a relationship are more likely to have sent sexually suggestive text messages ($\chi^2$ (1) = 4.86, $p < .05$; 39.1% of those in a relationship, 24.8% of those who were single), sent text messages propositioning sexual activity ($\chi^2$ (1) = 5.37, $p < .05$; 27.2% of those in a relationship, 14.2% of those who were single), sent self-produced sexually suggestive photos ($\chi^2$ (1) = 4.70, $p < .05$; 22.8% of those in a relationship, 11.5% of those who were single), and sent self-produced nude photos ($\chi^2$ (1) = 5.22, $p < .05$; 17.4% of those in a relationship, 0.1% of those who were single).

Next, further analyses were conducted on behaviours of sending sexts. It was found that majority of the participants had engaged in the sexting behaviours for more than five times. Additionally, across all four sexting behaviours, the recipient of the sext was reported to be mainly one's romantic partner, that is the boyfriend or the girlfriend.

Table 17.1. Descriptive statistics of sexting behaviours.

| | Sexually Suggestive Text Message (n = 64) | Text Message Propositioning Sexual Activity (n = 41) | Self-produced Sexually Suggestive Photo (n = 34) | Self-produced Nude Photo (n = 24) |
|---|---|---|---|---|
| **Frequency of sexting behaviours** | | | | |
| Once | 6 | 3 | 3 | 3 |
| Two Times | 8 | 3 | 4 | 5 |
| Three to Five Times | 11 | 11 | 8 | 4 |
| More than Five Times | 39 | 24 | 19 | 12 |
| **Target of sexting behaviours** | | | | |
| Boyfriend/Girlfriend | 54 | 36 | 30 | 22 |
| Someone I had a crush on | 6 | 4 | 4 | 3 |
| Someone I dated or hooked up with | 21 | 14 | 12 | 5 |
| Someone I wanted to date or hook up with | 12 | 12 | 6 | 4 |
| One or more good friends | 7 | 2 | 2 | 1 |
| Someone I only know online | 10 | 6 | 3 | 3 |
| Someone who forced or blackmailed me | 0 | 0 | 0 | 0 |
| Others | 2 | 1 | 1 | 0 |
| **Motivation of sexting behaviours** | | | | |
| To get the recipient's attention | 16 | 9 | 14 | 11 |
| To initiate sexual behaviour with the recipient | 30 | 29 | 17 | 14 |

| | | | | |
|---|---|---|---|---|
| As a "sexy" present to boyfriend/girlfriend | 24 | 14 | 20 | 18 |
| To feel sexy | 19 | 12 | 11 | 11 |
| As a joke | 24 | 12 | 6 | 3 |
| To be fun/flirtatious | 47 | 26 | 18 | 12 |
| In response to one that was sent to me | 38 | 15 | 17 | 12 |
| To enhance my relationship with the recipient | 28 | 22 | 15 | 14 |
| Because I was pressured | 2 | 0 | 5 | 3 |
| By force or blackmail | 0 | 0 | 0 | 0 |
| Others | 2 | 1 | 2 | 0 |
| **Medium used for sexting behaviours** | | | | |
| Traditional messaging (SMS; MMS) | 23 | 19 | 6 | 5 |
| Applications | 50 | 29 | 29 | 21 |
| Websites | 6 | 3 | 4 | 3 |
| Others | 3 | 3 | 2 | 2 |
| **Consequences of sexting behaviours** | | | | |
| No consequence | 16 | 8 | 7 | 4 |
| Positive consequence | 46 | 32 | 22 | 16 |
| Negative consequence | 17 | 14 | 12 | 10 |

An analysis of the motivations of the sexting data suggested that the top few reasons for sexting were: (1) to initiate sexual behaviour with the recipient; (2) to be fun or flirtatious; (3) to enhance their relationship with the recipient; (4) in response to a sext that was received; and (5) as a 'sexy' present to the boyfriend or the girlfriend. Sexting behaviours were reported to be typically conducted over mobile applications such as WhatsApp, Telegram, and Snapchat, where 70% to 87% of the participants who reported past sexting behaviours had done it over these applications. Lastly, more participants reported experiencing positive consequences (e.g., "Flirting increased self-esteem") as compared to negative consequences (e.g., "If discovered, people will judge me").

## 17.5 Implications

### 17.5.1 *Many reported receiving sexts more than sending one*

Results of this study showed that consistent with past international studies such as those conducted in the US [e.g., Aldridge *et al*., 2013], similar trends in sexting were found in Singapore. In this study, more participants reported having received sexts (22% to 43%) as compared to sending one (11% to 31%). This study also looked at four sexting behaviours with differing intensity and found that there is a decreasing trend in the frequency of occurrence where more participants reported engaging in sexting behaviour of lower intensity (i.e., sending sexually suggestive text messages) and less participants engaged in the higher intensity ones (e.g., sending self-produced nude photos).

Based on a survey of social morality conducted by the Institute of Policy Studies (IPS) in 2014, Singapore is a largely conservative society [Tham & Mokhtar, 2014]. Despite this, sexting is not an uncommon behaviour. In fact, this study found that 31% of the participants reported sending sexually suggestive text messages and 11.7% of the participants reported sending self-produced nude photos. This is similar to the study conducted by Strohmaier, Murphy, and DeMatteo [2014], which found that between 33% to 40% of teenagers and young adults have engaged in some form of sexting, while 15% to 20% sent semi-nude or nude pictures to others. Moreover, most participants reported having engaged in the sexting behaviour for more than five times.

### 17.5.2 *Sexting is more common and frequent than expected*

The earlier findings imply that sexting is more common and frequent than expected. This finding suggests support for the online disinhibition effect, where individuals are likely to do and say things in cyberspace that they will not usually do in other settings [Suler, 2014]. Characteristics of the internet, especially in applications such as WhatsApp and Snapchat in the case of sexting, reduced one's inhibition and thus, they may engage in sexting behaviours, even in the conservative Singapore society. This is also supported by the responses of participants, such as "My partner and I got closer because I could express things I wouldn't in real life."

### 17.5.3 *Sexting is common amongst those in a relationship*

In contrast to past findings that most individuals engaged in sexting due to peer pressure [Dake *et al*., 2012], this study found that sexting was mainly done in the context of relationships, where recipients tended to be one's boyfriend or girlfriend. Sexting was used as a medium to improve one's relationship with the recipient by initiating sexual behaviour, flirting, or as a present to the recipient. Interestingly, among those who have engaged in sexting behaviours, participants reported experiencing more positive consequences—such as increasing self-esteem and enhancing relationship with the recipient—as compared to negative ones. Although past research has mainly looked into the negative consequences of sexting, the findings of this study suggest that there are also positive consequences associated with it.

### 17.5.4 *Males are more likely to be perpetrators of sexting, etc.*

While some past studies found mixed findings with regards to gender differences in sexting [e.g., Klettke *et al*., 2014], this study found that males were more likely to have received sexually suggestive photos self-produced by the sender and, in turn, sent sexually suggestive text messages and text messages propositioning sexual activity. Additionally, individuals who were in a relationship reported more past behaviours of sending sexts as compared to individuals who were single. This is

consistent with the finding that most recipients of sext are the sender's girlfriend or boyfriend, and that sexting was done primarily to maintain and enhance relationship with the recipient.

## 17.6 Psychosocial Risk and Protective Factors of Sexting

The psychosocial perspective has been used to understand many behaviours, such as adolescent sexual behaviours [Santelli *et al*., 2004]. Psychosocial theories are useful as they incorporate both dispositional factors which are stable, and situational factors which are amendable to change [Teese & Bradley, 2008]. It can include personality, cognitive, affective variables, and also those reflecting the influence of family, peers and community. This section explores sexting from the psychosocial view by examining the various dispositional, cognitive, and social risk and protective factors that may influence one's intention to sext, and in turn looks at how these factors affect sexting behaviours.

### 17.6.1 *Dispositional factors related to sexting*

From research, four key dispositional factors were found to be correlated with sexting.

#### 17.6.1.1 *Histrionic personality*

A study conducted by Ferguson [2011] found that after controlling for age, histrionic traits could significantly predict sexting. Individuals with Histrionic Personality Disorder (HPD) adopt an interaction style that is seductive, emotionally shallow, and dramatic [Ferguson & Negy, 2014]. These individuals demonstrate traits such as motivation to seek the centre of attention, seductiveness, flirtatiousness, or being sexually provocative. These traits of histrionic personality were posited to be positively associated with sexting behaviours [Rosen *et al*., 2013].

#### 17.6.1.2 *Sensation seeking*

Sensation seeking refers to the tendency to seek out new and exciting experiences [Dir *et al*., 2013]. Individuals who ranked high on sensation

seeking tendencies tended to be more sensitive to rewards than punishment cues. These individuals are more influenced by the short-term benefits of a behaviour than compared to its long-term costs. As a result, they were more likely to engage in risky behaviours to seek or enhance pleasure, despite the potential negative consequences. For example, sensation seeking has been found to be associated with alcohol and drug use, and risky sexual behaviours [Cooper *et al*., 2003]. In the context of sexting, past studies found significant positive association between sensation seeking and sexting behaviours [Gomez & Ayala, 2014; Ouytsel *et al*., 2014].

#### 17.6.1.3 *Attachment anxiety*

One's attachment style is said to set the stage for their future social interaction with others [Weisskirch & Delevi, 2011], and this could in turn affect one's likelihood of engaging in sexting. Past research on sexting has examined the association with attachment styles from two dimensions: (1) attachment avoidance and (2) attachment anxiety. While mixed findings were found between attachment avoidance and sexting, attachment anxiety was often found to be associated with sexting [Drouin & Landgraff, 2012; Weisskirch *et al*., 2016]. Anxiously attached individuals tended to experience a strong desire for closeness, but also an intense fear of abandonment or separation at the same time [Drouin & Landgraff, 2012]. Previous research found that these individuals engaged in unwanted sex in order to reduce relational insecurity in the hopes of maintaining their partner's interest in the relationship [Weisskirch & Delevi, 2011]. Thus, sexting may be another form of reassurance-seeking behaviour for individuals who score high on attachment anxiety.

#### 17.6.1.4 *Religiosity*

Religiosity affects one's behaviour as it influences one's self-control and delay of self-gratification [Hall *et al*., 2016]. Past research found it to be related to a variety of risky behaviours such as alcohol use and drug abuse [Mohammadpoorasl *et al*., 2014]. Religiosity has also been found to affect the frequency of one's sexual behaviours; individuals who are more religious reported lower frequency of sexual behaviours [Penhollow *et al*., 2005]. Furthermore, Strassberg, Rullo, and Mackaronis [2014] found

religiosity to be negatively associated with both sending and receiving sexts. Thus, religiosity is proposed to be a protective factor against sexting behaviours.

### 17.6.2 *Cognitive factors related to sexting*

There are three cognitive factors of interest.

#### 17.6.2.1 *Attitudes toward sexting*

Based on a systematic literature review, Klettke, Hallford, and Mellor [2014] found a positive relationship between having a favourable attitude toward sexting and engagement in sexting behaviour. Consistent with this finding, Hudson [2011] reported a trend where teenagers and young adults who have a more positive attitude toward sexting were more likely to engage in it. Conversely, it is logical to postulate that individuals with negative attitudes toward sexting would be less likely to engage in sexting.

#### 17.6.2.2 *Resistance to peer pressure*

As with other risky behaviours, peer pressure plays an important role in sexting. Past research [e.g., Ouytsel *et al*., 2014] found that young people were more influenced by the perceived social pressure from their social referents, especially their peers and romantic partner, as compared to their evaluation of the possible consequences associated with sexting behaviours. Peer pressure has been found to be one of the main reasons people engage in sexting; 23% of teens reported that they sexted due to pressure from a friend, and 51% of teenage girls felt pressured by a boy to send sexually explicit messages [Dake *et al*., 2012]. Thus, one's ability to resist to peer pressure may act as a protective factor against engagement of sexting.

#### 17.6.2.3 *Knowledge of sexting*

Education has been found to be an essential factor in preventing sexting [Aldrige *et al*., 2013]. Education can help individuals to recognise

the long-term social, emotional, and legal consequences of sexting. Thus, knowledge of sexting may act as a protective factor which may lower the likelihood of one's engagement in sexting. A study by Strohmaier, Murphy, and Dematteo [2014] found that having an awareness of the legal consequences of sexting was negatively associated with sexting behaviours as minors. Furthermore, participants reported that they would be deterred from sexting if they were aware of the consequences.

## 17.7 Social Factors

There are two social factors of interest.

### 17.7.1 *Perceived subjective norms*

Perceived subjective norms have been one of the key determinants of behaviours, and this perception can be affected by how prevalent individuals think the behaviour is, and whether they know of peers who engaged in certain kinds of behaviours [Ajzen, 1991]. A study conducted by Hudson [2011] found subjective norms toward sexting was the strongest predictor of whether someone has engaged in sexting. Additionally, Lippman, and Campbell [2012] found that adolescents' perceived norms regarding the frequency of sexting behaviours were consistent with their reported behaviours of sexting if they believed that it was a common behaviour among their peers. Thus, individuals with perceived subjective norms that support sexting are more likely to engage in such behaviour.

### 17.7.2 *Family support*

Aldridge, Arndt, and Davies [2013] found that students who had supportive families were less likely to participate in sexting. Family forms a huge part of most people's lives, and a strong familial support system has been found to be a protective factor against many risky behaviours. For instance, Mohammadpoorasl, Ghahramanloo, Allahverdipour, and Augner [2014] found adolescents from cohesive families were less likely to engage in alcohol or drug use. Hence, it is logical to postulate that family support may also act as a protective factor against sexting behaviours

## 17.8 Psychosocial Framework for Sexting: Exploring the Relationship Between Intention and Behaviours for Sexting

One of the key theories that looked into the relationship between one's intention and behaviour is the Theory of Planned Behaviour (TPB). According to the TPB developed by Ajzen [1991], one's behaviour is directly determined by one's intention, which indicates how much effort one is willing to exert to perform that behaviour. Thus, it is postulated that those who have a strong intention to sext would be more likely to engage in sexting behaviours. This is also supported by Walrave, Heirman, and Hallam [2014], who found a significant positive association between the intention of sexting and self-reported sexting behaviours. Thus, this section also aims to examine the relationship between one's intention to sext and actual sexting behaviours[b] using the abovementioned nine psychosocial variables (see Figure 17.1).

As there are many variables explored in the study, a three-step approach[c] was used to analyse the relationship between the risk and protective factors and sexting. First, correlation was computed to examine the relationship between all the psychosocial factors, intention to sext, and sexting behaviours. Next, regression was conducted to investigate the relationship between these variables and one's intention to sext. Lastly, Structural Equation Modelling (SEM) was used to assess the relationship between the psychosocial factors, intention to sext, and sexting behaviours altogether.

Based on the findings from SEM, four factors were found to be key predictors of four types of sexting behaviours (i.e., sending sexually suggestive text message, sending text message propositioning sexual activity and sending self-produced sexually suggestive photo). They are 'intention to sext', 'attitudes (positive) toward sexting', 'histrionic personality', and 'sensation seeking'.

---

[b] Participants were required to report both past receiving and sending behaviours of sexting. The participants' frequency, target audience, motivation, medium used and consequences of their sexting behaviours in an attempt to better understand sexting behaviours.

[c] For more information about the analysis, please contact the authors.

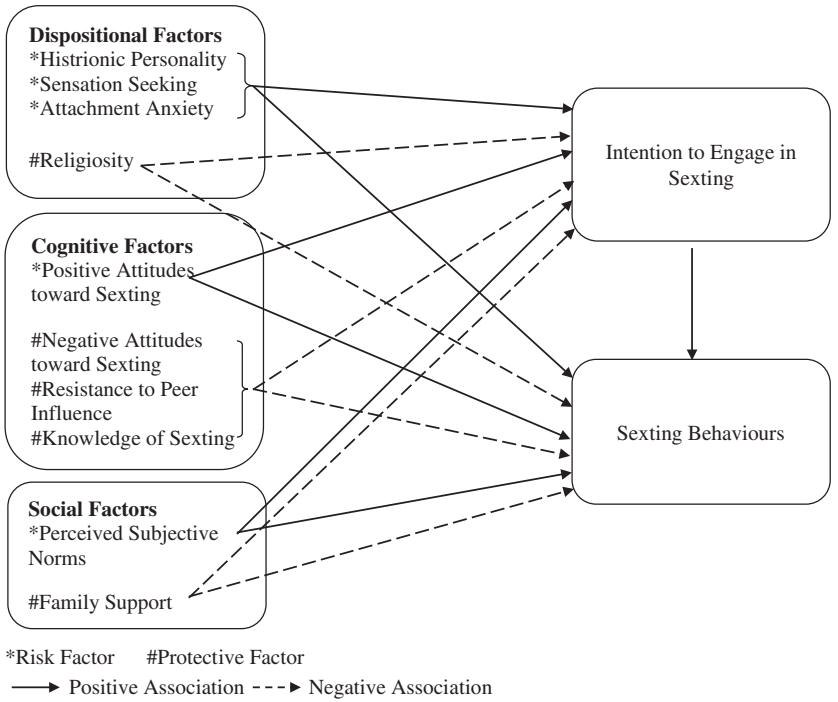



Figure 17.1. Summary of variables being investigated.

### 17.8.1 *Intention to sext*

'Intention to sext' was found to be a significant predictor of all four sexting behaviours, where greater intention is associated with a higher likelihood of having engaged in the behaviours. This finding is consistent with the Theory of Planned Behaviour (TPB) developed by Ajzen [1991] which postulated that one's behaviour is directly determined by one's intention. Walrave, Heirman, and Hallam [2014] also found that with the exception of behaviours that are not within one's control, intention was the strongest predictor for actual or self-reported behaviour, and this was also supported by the results of this study.

### 17.8.2 *Attitudes (positive) towards sexting*

'Attitudes towards sexting' was found to be the strongest predictor of 'intention to sext', where positive attitudes toward sexting was associated

with a greater intention to sext. It also significantly predicted one's engagement in sending sexually suggestive text messages, text messages propositioning sexual activity, and self-produced sexually suggestive photos. This provided support for the theory that one's attitude will predict his behaviours due to the inner need to maintain consistency and avoid dissonance [Ajzen & Fishbein, 1977]. Thus, positive attitudes toward sexting was an important risk factor that may not only predict one's intention to sext, but also the actual behaviour.

### 17.8.3 *Histrionic personality* (*seductiveness*)

'Histrionic personality (seductiveness)' was found to be positively associated with 'intention to sext'. This implied that individuals who scored high on the seductive factor of histrionic personality were more likely to report a greater intention in sexting. These individuals may have turned to sexting as a means to appeal to their recipients. This coincided with the finding that most participants reported sexting with the intention to initiate sexual behaviours with the recipient.

### 17.8.4 *'Sensation seeking'* (*intensity*)

'Sensation seeking (intensity)' was found to be positively associated with 'intention to sext'. Individuals who scored higher on the intensity factor of sensation seeking tended to report a greater intention to sext, which is further supported by the finding that most participants reported having sexted more than five times. Thus, these individuals viewed sexting as an exciting activity, and engaged in it repeatedly despite the possible negative consequences.

## 17.9 Implications

The results of this study add theoretical value to the scarce existing knowledge of sexting behaviours. Psychosocial theory has been used to explain many behaviours, and sexting is no exception. The findings suggest that a combination of psychosocial factors can increase one's intention to sext, which then affects one's engagement in the behaviour. Specifically, this study found that dispositional traits might predispose one to be motivated to engage in seductive and risky behaviours, while

positive attitudes may lead one to engage in sexting to achieve such goals. Additionally, this is the first empirical study[d] that looks into sexting behaviours in Singapore. The results of this study imply that sexting is not uncommon in the local setting. Even though individuals who engage in sexting may enhance one's self-esteem and improve relationship with the sext's recipient, it is still potentially harmful and can lead to negative repercussions such as harassment and suicide. Thus, more attention should be given to look into this potentially harmful phenomenon.

## 17.10 Conclusion

In conclusion, this chapter attempts to build upon the scarce existing research on the phenomenon of sexting. Some main findings were noted. First, this study found that sexting is not uncommon behaviour in Singapore, where 55% of the participants reported having received at least one sext, and 35% reported having engaged in some form of sexting. Second, analysis revealed that most participants sexted in order to develop or enhance a relationship with the recipient, and that there were both positive and negative consequences associated with the act of sexting. Third, dispositional traits, particularly histrionic personality and sensation seeking, can predispose one to have a stronger intention to sext. Additionally, having positive attitudes toward sexting were also associated with the engagement of sexting behaviours. As sexting can lead to negative repercussions, more attention should be given to looking into this phenomenon. Thus, future research can look into the dynamics of sexting behaviours in teenagers, and also, explore some other possible protective factors in order to better tackle this potentially harmful phenomenon.

---

[d]One of the methodological constraints of this study is the use of convenience sampling. Furthermore, as most of the participants are university graduates, this limits the generalizability of the findings to other population such as adolescents. Thus, future research can make use of random or even stratified sampling in order to increase the external validity of their study. Additionally, adolescents may be more vulnerable to sexting and thus, future research should also look into the phenomenon of sexting in Singapore adolescents and investigate what are the possible risk and protective factors of sexting for them.

## 17.11 Acknowledgement

The views expressed in this chapter are the authors' alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 17.12 References

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, *50*, pp. 179–211.

Ajzen, I., & Fishbein, M. (1977). Attitude-behaviour relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), pp. 888–918.

Aldridge, M. J., Arndt, K. J., & Davies, S. C. (2013). Sexting: You found the sext, what to do next? How school psychologists can assist with policy, prevention, and intervention. *The Ohio School Psychologist*, *58*(2), pp. 6–10.

Aldridge, M. J., Davies, S. C., & Arndt, K. J. (2013). Is your school prepared for a sexting crisis? *Principal Leadership*, *13*(9), pp. 12–16.

Arnett, J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, *16*(2), pp. 289–296.

Attorney-General's Chamber. (2015). Singapore Statues Online. Retrieved from http://statutes.agc.gov.sg/aol/search/display/view.w3p;ident=bfddb376-a7b9-4597-b7df-2050d6735826;page=0;query=DocId%3A%22025e7646-947b-462c-b557-60aa55dc7b42%22%20Status%3Ainforce%20Depth%3A0;rec=0#pr299-he-

Cooper, M. L., Wood, P. L., & Orcutt, H. K. (2003). Personality and the predisposition to engage in risky or problem behaviours during adolescence. *Journal of Personality and Social Psychology*, *84*(2), pp. 390–410.

Crimmins, D. M., & Seigfried-Spellar, K. C. (2014). Peer attachment, sexual experiences, and risky online behaviour as predictors of sexting behaviours among undergraduate students. *Computers in Human Behaviour*, *32*, pp. 268–275.

Dake, J. A., Price, J. H., & Maziarz, L. (2012). Prevalence and correlates of sexting behaviour in adolescents. *American Journal of Sexuality Education*, *7*, pp. 1–15.

Dir, A. L., Cyders, M. A., & Coskunpinar, A. (2013). From the bar to the bed via mobile phone: A first test of the role of problematic alcohol use, sexting, and impulsivity-related traits in sexual hookups. *Computers in Human Behaviour*, *29*, pp. 1,664–1,670.

Doring, N. (2014). Consensual sexting among adolescents: Risk prevention through abstinence education or safer sexting? *Journal of Psychosocial Research on Cyberspace*, *8*(1).

Drouin, M., & Landgraff, C. (2012). Texting, sexting, and attachment in college students' romantic relationships. *Computers in Human Behaviour*, *28*, pp. 444–449.

Ferguson, C. J. (2011). Sexting behaviours among young Hispanic women: Incidence and association with other high-risk sexual behaviours. *Psychiatric Quarterly*, *82*, pp. 239–243.

Ferguson, C. J., & Negy, C. (2014). Development of a brief screening questionnaire for histrionic personality symptoms. *Personality and Individual Differences*, *66*, pp. 124–127.

Fichten, C. S., Nguyen, M. N., Amsel, R., Jorgensen, S., Budd, J., Jorgensen, M., Asuncion, J., & Barile, M. (2014). How well does the Theory of Planned Behaviour predict graduation among college and university students with disabilities? *Social Psychology of Education*, *17*(4), pp. 657–685.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item-response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*, pp. 350–365.

Glock, C. Y. (1962). On the study of religious commitment. *Religious Education: The official journal of the Religious Education Association*, *57*(4), pp. 98–110.

Gomez, L. C., & Ayala, E. S. (2014). Psychological aspects, attitudes and behaviour related to the practice of sexting: A systematic review of the existent literature. *Procedia—Social and Behavioural Sciences*, *132*, pp. 114–120.

Hall, M., Williams, R. D., Ford, M. A., Cromeans, E. M., & Bergman, R. J. (2016). Hooking-up, religiosity, and sexting among college students. *Journal of Religion and Health*, pp. 1–13.

Hoge, R. (1972). A validated intrinsic religious motivation scale. *Journal for the Scientific Study of Religion*, *11*(4), pp. 369–376.

Hudson, H. K. (2011). Factors affecting sexting behaviours among selected undergraduate students (Unpublished doctoral dissertation). Southern Illinois University, Carbondale.

Klettke, B., Hallford, D. J., & Mellor, D. J. (2014). Sexting prevalence and correlates: A systematic literature review. *Clinical Psychology Review*, *34*, pp. 44–53.

Korenis, P., & Billick, S. B. (2014). Forensic implications: Adolescent sexting and cyberbullying. *Psychiatric Quarterly*, *85*(1), pp. 97–101.

Lee, C. H., Moak, S., & Walker, J. T. (2016). Effects of self-control, social control and social learning on sexting behavior among south Korean youths. *Youth & Society*, *48*(2), pp. 242–264.

Lippman, J. R., & Campbell, S. W. (2012). Teenagers and sexting: Perceived norms and sexual double standard. Paper presented at the ICA Conference, 24–28 May 2012, Phoenix, AZ.

Messer, D. G., Bauermeister, J. A., Grodzinski, A., & Zimmerman, M. (2013). Sexting among young adults. *Journal of Adolescent Health*, *52*(3), pp. 301–306.

Mohammadpoorasl, A., Ghahramanloo, A. A., Allahverdipour, H., & Augner, C. (2014). Substance abuse in relation to religiosity and familial support in Iranian college student. *Asian Journal of Psychiatry*, 9, pp. 41–44.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Outysel, J. V., Gool, E. V., Ponnet, K., & Walrave, M. (2014). Brief report: The association between adolescents' characteristics and engagement in sexting. *Journal of Adolescence*, *37*, pp. 1,387–1,391.

Ouytsel, J. V., Walrave, M., & Gool, E. V. (2014). Sexting: Between thrill and fear—How schools can respond. *The Clearing House*, *87*, pp. 204–212.

Penhollow, T., Young, M., & Bailey, W. (2007). Relationship between religiosity and "hooking up" behaviour. *Journal of Health Education*, *36*(2), pp. 75–85.

Procidano, M. E., & Heller, K. (1983). Measures of perceived social support from friends and from family: Three validation studies. *American Journal of Community*, *11*(1), pp. 1–24.

Rosen, L. D., Whaling, K., Rab, S., Carrier, L. M., & Cheever, N. A. (2013). Is Facebook creating "iDisorders"? The link between clinical symptoms of psychiatric disorders and technology use, attitudes and anxiety. *Computers in Human Behaviour*, *29*(3), pp. 1,243–1,254.

Santelli, J. S., Kaiser, J., Hirsch, L., Radosh, A, Simkin, L., & Middlestadt, S. (2004). Initaitla of sexual intercourse among middle school adolescents: The influence of psychosocial factors. *Journal of Adolescent Health*, *34*, pp. 200–208.

Steinberg, L., & Monahan, K. C. (2007). Age differences in resistance to peer influence. *Developmental Psychology*, *43*(6), pp. 1,531–1,543.

Stober, J. (2001). The social desirability scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, *17*, pp. 222–232.

Strassberg, D. S., Rullo J. E., & Mackaronis, J. E. (2014). The sending and receiving of sexually explicit cell phone photos ("Sexting") while in high school: One college's students' retrospective reports. *Computers in Human Behaviour*, *41*, pp. 177–183.

Strohmaier, H., Murphy, M., & DeMatteo, D. (2014). Youth sexting: Prevalence rates, driving motivations, and the deterrent effect of legal consequences. *Sexuality Research and Social Policy*, *11*, pp. 245–255.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behaviour*, *7*(3), pp. 321–326.

Tai, J. (2015). The dangers in the rise of 'sexting' among teenagers in Singapore. *The Straits Times*. Retrieved from http://www.straitstimes.com/singapore/the-dangers-in-rise-of-sexting-among-teenagers-in-singapore

Teese, R., & Bradley, G. (2008). Predicting recklessness in emerging adults: A test of a psychosocial model. *The Journal of Social Psychology*, *148*(1), pp. 105–126.

Teng, A. (2015). More secondary students involved in 'sexting': Poll. *The Straits Times*. Retrieved from http://www.straitstimes.com/singapore/more-secondary-students-involved-in-sexting-poll

Texas School Safety Center. (n.d.). Before you text Sexting Prevention Educational Program. Retrieved from https://txssc.txstate.edu/tools/courses/before-you-text/

Tham, Y. C., & Mokhtar, M. (2014). Singaporeans still largely conservative, IPS survey finds. *The Straits Times*. Retrieved from http://www.straitstimes.com/singapore/singaporeans-still-largely-conservative-ips-survey-finds

Walrave, M., Heirman, W., & Hallam, L. (2014). Under pressure to sext? Applying the theory of planned behaviour to adolescent sexting. *Behaviour & Information Technology*, *33*(1), pp. 86–98.

Weisskirch, R. S., & Delevi, R. (2011). "Sexting" and adult romantic attachment. *Computers in Human Behaviour*, *27*, pp. 1,697–1,701.

Weisskirch, R. S., Drouin, M., & Delevi, R. (2016). Relational anxiety and sexting. The *Journal of Sex Research*, *1*, pp. 1–9.

This page intentionally left blank

# Part 4

# Conclusion

This page intentionally left blank

https://doi.org/10.1142/9789811232411_0018

# Chapter 18

# The Future of Cyber-Forensic Psychology: How to Prepare

Majeed Khader

*Home Team Behavioural Sciences Centre,*
*Ministry of Home Affairs, Singapore*

*Khader_MAJEED@mha.gov.sg*

> Forensic science is the study of the physical evidence at a crime scene—fibers or bodily fluids or fingerprints. In TV terms, think CSI. Forensic psychology is the study of the behavioural evidence left behind at the crime scene, what we like to call 'the blood spatter of the mind'. Then there's my area, forensic cyberpsychology, which focuses on the cyberbehavioral evidence … Every contact leaves a trace … This is just as true in cyberspace.
>
> (Aiken, 2016, p. 6)

## 18.1 Introduction

This final chapter discusses two issues. First, what we can expect for the future for cyber forensic psychology, and second, how we should respond.

On the first issue, we must foremost expect that human criminal needs will not change in the future. They have not changed since the dawn of human civilisations and will not in the future. The physical manifestations of future crime may look different, but their essence will remain. Criminals will continue to be driven by greed, lust, envy, and the usual

criminogenic needs. Crime will also exist for the same psychological, sociological, and economic reasons that have always influenced criminality. However, the cyberspace will change their external manifestations, making criminal victimisation more accessible, real-time, sophisticated, harder to track and trace, and worryingly more intimate.

In Fattah (1997)'s discussion on the future of criminology and criminology of the future, he noted Tom Keenan (1984)'s prediction about how technology could facilitate new crimes:

> Criminals of the future will probably have many of the same motives as today's crooks … greed, lust, and revenge … but the ways in which they carry out their crimes may be radically different. With technology now into development, a criminal will be able to invade your home using computer links, telephone and two-way video taps. He or she may attack you with psychological harassment or mind manipulation techniques, demanding protection money to stay out of your brain. New technology will make all these things possible, even likely. (cited in Fattah, 1997, p. 284)

Another consistent and astute perspective comes from the world's best (ex-)con-man, Frank Abagnale, who specialised in impersonation and forgery. Frank was played by Leonardo DiCaprio in the 2002 film 'Catch Me If You Can'. At the age of 16, Frank posed as a pilot for Pan American World Airways in order to wangle free flights across the world. He also pretended to be a doctor, before masquerading as an attorney. After his 12-year sentence, he was paroled on the condition that he helped the FBI. He has since made a career as a security consultant, working closely with the FBI for 40 years. Frank described the ease of committing fraud, and further highlighted how cybercrime could be facilitated in the age of disruption:

> What I did was almost 50 years ago and it's about 4,000 times easier today to con people than when I did it. To forge a cheque 50 years ago, you needed a Heidelberg printed press, you had to be a skilled printer, know how to do colour separations, negatives, typesetting … those presses were 90-feet long and 18-feet high. There was a lot of work involved in creating a cheque. Today, you open a laptop. If you are going to forge a British Airways cheque, you go to their website, capture the corporate logo and put it in the top right corner. You then put a jet taking

> off in the background and make a really fancy four-colour cheque in 15 minutes on your computer. You then go down to an office supply store, buy security cheque paper and put it in your colour printer. (Solon, 2017, paras. 4–5)

While people in our societies mature physically, this is not true in the digital sense. Chronological maturity does not commensurate with digital maturity. What we see is digital naivety. In contrast, cyber evolution has seen an unprecedented 'criminal evolution' (Henson *et al*., 2016). Criminals exploit the technological advancements in various ways as described within this book. Many examples such as live-streamed crimes, scams, hacking, robotics, sexting, online self-harm, and more, were discussed at length in the earlier chapters. We are already seeing a surge in cybercrimes globally and locally. In Singapore, traditional crimes are extremely low in numbers, despite the high urban density and large population we have in the city state; yet, cybercrimes are on the constant rise.

## 18.2 Recommendations

What can we do? The following are some possibilities in the form of strategy counterresponse.

### 18.2.1 *Make the building of cyber resilience in our communities a priority*

We need to train our communities to be more sophisticated about cybercrime and security issues. The more we drive on the cyber-tech highway, the more we need to appreciate that our communities require more preparation against threats in cyber and cyber-forensic areas. Singapore's Cyber Security 2016 document maps out excellent policy, legislative, regional capability building strategies, and calls for greater collective actions (CSA: Singapore's Cyber Security Strategy, 2016). However, if we recognise that the human potentially remains the weakest link in the chain, then more should be done to build 'people and community cyber resilience'. People must learn how to protect themselves.

### 18.2.2 *Governments must invest in cyber-forensic psychology and psychologists*

Governments and law enforcement agencies must not just invest in technological and forensic departments; they must also develop the human psychological side of things. They should invest in research on cyber forensic psychology, and employ more psychologists to tackle the emerging cyber threat. More is needed locally, even though Singapore has made a great start. The Cybersecurity Strategy carries four pillars: (1) building a resilient infrastructure to strengthen the critical infrastructures by working closely with private sectors and cyber security community; (2) creating a safer cyberspace by promoting involvement from not only government but also industry and the public; (3) developing a vibrant security ecosystem by working with industry and academia to grow the cyber security workforce; (4) strengthening international partnerships, especially among ASEAN members, to address transnational cyber security issues. We could also add a fifth pillar; and (5) grow a cadre of cyber-forensic psychologists to work hand in hand with law enforcement and security professionals. A better understanding of the psyche of the perpetrators and the victims as well as the mindsets of security professionals would propel Singapore into an even stronger position to manage this growing threat.

### 18.2.3 *Psychology departments must teach cyber psychology*

Psychology departments in universities need to employ professors and faculty with a background in technology, cyber forensic psychology, and cyber psychology. In the post COVID-19 era, it makes sense for psychology departments to invest strongly in 'Psych-Tech'. Not only should this be taught as a core module in all psychology curriculum in the same way that fundamentals of research design are taught, students should also receive instruction on how to design research studies and experiments compatible for mobile phone devices, robotics, machine learning, social media app development, virtual reality (VR) and augmented reality (AR) simulation, gamification, and wearables.

Psychology as a discipline appears to be slow in the uptake of cyber and technology worldwide, when compared to other disciplines such as

engineering and medicine. Professor Mary Aiken is a good example of psychology specialists who work in this field. As Director of the RCSI CyberPsychology Research Centre and Academic Advisor (Psychology) to the European Cyber Crime Centre (EC3) at Europol, Professor Aiken is not only an academic but also a practitioner. She is directly involved in various aspects of cybersecurity work, including child online safety, youth pathways into cybercrime, cyber criminology, human trafficking, as well as safety-tech and online safety technology (Aiken, 2016). Another equally important role for psychologists is to prevent the misuse of technology, social media, and e-data in ways which can harm humans. For example, there is much concern about how technology and social media have replaced the real-world analogue of relationships, communications, and carry privacy concerns. Psychologists should explore this area and examine the human vulnerability in the cyberspace.

### 18.2.4 *Crime prevention departments must appreciate the cyber frame and adopt a calibrated outreach strategy*

Today, much of crime prevention efforts appear to take on the form of traditional hard copy posters, television advertisements, and we see them on the walls of apartments, commercial buildings, and other physical premises. This may still be useful for the native natives (or perhaps even the older) population who are more familiar with such means of traditional outreach. Studies in Canada, for example, found that younger and more educated citizens are more satisfied with the use of social media in crime prevention efforts (Jones, 2013). Crime prevention policy advocates and law enforcement agencies, including police and drug prevention officers, would benefit in specialised training in social media analytics, social media marketing outreach, video production, and the gamut of other tools which enhance communication. One key takeaway in this area seems to be that social media crime prevention campaigning should be planned and target demographic groups who are likely to receive the intended message. The implications are significant. Senior citizens may not have interest in cyber campaigns, and therefore may be more easily targeted by criminals. The pressing operational agenda for law enforcement and security professionals is this: ‘when it comes to cyber-crime prevention, what works, when, for whom, and in what way?’

### 18.2.5 *We need to refine traditional criminology theories*

How have criminology theories responded to this evolution on cyber criminal behaviour? In thinking about situational crime prevention, it is true that we now have to think of crime within the 'cyber environment' — incorporating cyberspace mediums such as search engines, websites, blogs, and social media sites. In addition, we have to rethink the definition of crime and the respective criminology theories. Crime is typically defined in some part by local legislation as local law often defines legality; but what if crimes are transnational and global? What new laws and regulations are needed? Also, if this new environment constitutes mobile phones, communication devices, the Internet of Things (IoT), wearables, and Home Cloud devices; how does this affect criminological understanding? How do we recognise cyber bulling, cyber extortion, and online radicalisation in these private and personal modalities? A deeper study of cyber-forensic psychology and criminology may be the way forward to address these issues.

### 18.2.6 *Build private-public partnerships to fight cybercrime*

Finally, it is undoubtedly important for both public and private companies to work together to combat cybercrime collectively. Should governing agencies talk to technology companies and social media companies to build security in their product design? Yes, it has to be so. After all, a secure product will increase customer satisfaction, so it makes good business sense for the private businesses to build good private-public partnerships to fight cybercrime. In Singapore, it is therefore enlightening to see that the Singapore government Cyber strategy does involves partnerships with the private sector.

## 18.3 Acknowledgement

The views expressed in this chapter are the author's alone and do not represent the official position or view of the Ministry of Home Affairs, Singapore.

## 18.4 References

Aiken, M. (2016). *The cyber effect: A pioneering cyber-psychologist explains how human behavior changes online*. Spiegel & Grau.

CSA: Singapore's Cyber Security Strategy. (2016, October 10). https://www.csa.gov.sg/news/publications/singapore-cybersecurity-strategy

Fattah, E. A. (1997). *Criminology: Past, present and future: A critical overview*. Springer.

Henson, B., Reyns, B. W., & Fisher, B. S. (2016). *Cybercrime victimization.* In C. A. Cuevas & C. M. Rennison (Eds.), *The Wiley handbook on the psychology of violence* (pp. 555–570). Wiley Blackwell. https://doi.org/10.1002/9781118303092.ch28

Keenan, T. (1984). *Crimes of the future*. Transcript of CBC program IDEAS (Oct. 15–29). Canadian Broadcasting Corporation.

Solon, O. (2017, October 04). Frank Abagnale on the death of the con artist and the rise of cybercrime. *Wired*. https://www.wired.co.uk/article/frank-abagnale