De Gruyter Handbook of Digital Criminology

De Gruyter Contemporary Social Sciences Handbooks

Volume 6

De Gruyter Handbook of Digital Criminology

Edited by
Mareile Kaufmann and Heidi Mork Lomell

DE GRUYTER

ISBN 978-3-11-106193-1
e-ISBN (PDF) 978-3-11-106203-7
e-ISBN (EPUB) 978-3-11-106264-8
ISSN 2747-9269
DOI https://doi.org/10.1515/9783111062037

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For details go to https://creativecommons.org/licenses/by-nc-nd/4.0/.

Creative Commons license terms for re-use do not apply to any content (such as graphs, figures, photos, excerpts, etc.) not original to the Open Access publication and further permission may be required from the rights holder. The obligation to research and clear permission lies solely with the party re-using the material.

Library of Congress Control Number: 2024950681

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

© 2025 with the author(s), editing © 2025 Mareile Kaufmann and Heidi Mork Lomell, published by Walter de Gruyter GmbH, Berlin/Boston, Genthiner Straße 13, 10785 Berlin.
This book is published with open access at www.degruyter.com.

Cover image: silverjohn / iStock / Getty Images Plus
Printing and binding: CPI books GmbH, Leck

www.degruyter.com
Questions about General Product Safety Regulation:
productsafety@degruyterbrill.com

Table of Content

Mareile Kaufmann and Heidi Mork Lomell
An introduction to digital criminology —— 1

Anne-Marie McAlinden
1 Abuse —— 21

Mareile Kaufmann
2 Accessing online communities —— 29

Mark A. Wood and Will Arpke-Wales
3 Affordances —— 37

Susanne Krasmann
4 Agency —— 45

Matthias Leese
5 Algorithm —— 53

Silje Bakken
6 App-based textual interviews —— 61

Nanna Bonde Thylstrup and Kristin Veel
7 Archives —— 69

Heather Dewey-Hagborg
8 Art as method —— 79

Rosamunde Van Brakel
9 Artificial intelligence —— 83

Monique Mann
10 Automation —— 91

Marion Oswald and Angela Paul
11 Bias —— 99

Aleš Završnik
12 Big data —— 107

Ravida Din and Shoshana Magnet
13 Biometric failure — 115

Julien Jeandesboz
14 Borders and border control — 123

Katja Franko
15 Categorization and sorting — 133

Stefano Mazzilli Daechsel
16 Computation — 141

Thomas J. Holt and Karen M. Holt
17 Cybercrime — 149

Meropi Tzanetakis
18 Darknet — 157

Joanna Redden
19 Data justice — 165

Rocco Bellanova
20 Databases — 173

Janet Chan
21 Datafication — 181

Daniel Trottier
22 Digilantism — 189

Jacqueline D. Wernimont
23 Digital — 197

Mareile Kaufmann
24 DNA / big genome data — 205

Claudia Aradau
25 Error — 215

Annette Markham
26 Ethics — 223

Pete Fussey
27 Facial recognition — 233

Anthony Amicelle
28 Financial crime and surveillance — 243

David S. Wall
29 Hacking — 251

Robin Cameron, Gregory Stratton, and Anastasia Powell
30 Hate crime and networked hate — 261

Malcolm Langford, Amelia Svensson, and Tone Wærstad
31 Identify theft — 269

Miria Grisot and Elena Parmiggiani
32 Infrastructures — 279

Helene Oppen Ingebrigtsen Gundhus and Jenny Maria Lundgaard
33 Intelligence — 287

Sanja Milivojevic
34 Internet of things — 297

Catherine Adams and Terrie Lynn Thompson
35 Interviews with digital objects — 305

Stefano Mazzilli Daechsel
36 Labs — 315

Maja Vestad
37 Low-tech — 323

Avital Mentovich and Orna Rabinovich Einy
38 Online courts — 333

Nicholas Gibbs and Alexandra Hall
39 Online ethnography — 343

Simon Egbert
40 Platforms — 353

Dean Wilson
41 Policing — 363

Emils Kilis, Helene O. I. Gundhus, and Vasilis Galis
42 Prediction — 371

Lee A. Bygrave
43 Privacy and data protection — 381

Heidi Mork Lomell
44 Privatization — 389

Rose Lunde and Peter Scharff Smith
45 Punishment — 399

Jan Christoffer Andersen
46 Recruitment via social media — 409

Maryia Šupa
47 Researching online forums — 419

Naomi Lintvedt and Kristin Bergtora Sandvik
48 Robots — 427

Pamela Ugwudike
49 Sentencing and risk assessment algorithms — 435

Helen M. Rand
50 Sex work — 445

Keith Hayward
51 Smart city — 453

Mirjam Abigail Twigt
52 Social media — 461

David Lyon
53 Surveillance — 469

Katja de Vries
54 Synthetic data and generative machine learning — 483

Dana Wilson-Kovacs
55 Translation —— 493

Sandra Walklate
56 Victimization —— 501

Sofia Ranchordas and Malou Beck
57 Vulnerability —— 509

List of contributors —— 519

Mareile Kaufmann and Heidi Mork Lomell

An introduction to digital criminology

"The Internet will disappear," is a famous sentence uttered by Eric Schmidt, then-Google CEO visiting the 2015 World Economic Forum (Smith 2015, online). His vision refers to the ways in which one day the devices surrounding us will be seamlessly connected, rendering our online existence from a deliberate choice into a default state. Indeed, being online is an everyday experience for many, where it tends to slip into the "technological unconscious" (Thrift, 2004). The same applies to our use of offline digital devices: it is easy to forget about their presence and to reflect about the ways in which they influence our everyday lives. Yet, even if unconscious, the internet and other digital infrastructures do not *disappear.* Not only can they be unintuitive, break down, or act back in unforeseen ways, disrupting our supposedly seamless experience. But their material existence keeps confronting us with societal and political issues. They raise questions about ownership (who owns infrastructures and data?), design (which values and functionalities are embedded in the design of digital devices?), practice (how do technologies enable new forms of usage?), choice (which choices can we take ourselves, which are taken for us?), ethics, governance, and regulation (what power structures and dilemmas does digitalization entail?).

This book pays attention to the ways in which electronic digital devices, online and offline, spread and cut across all fields of criminology. Criminologists need to be able to identify, make visible and analyze the technologies, practices, and social dynamics that emerge from their growing relevance. Digitalization changes interaction and information flow, speed and spatiality, experiences and practices of crime and crime control. It influences every aspect of criminology, such as victimization and violence, corporate and organized crime, fraud, drug markets and consumption, radicalization and terrorism, policing and forensics, sentencing and punishment. Digitalization bears new modi operandi, cultures of surveillance and social control, and new types of offenses, which this book documents and problematizes.

Hence, this is a handbook about the *digitalization* of criminological arenas rather than criminology in the 'digital era.' Criminology has not entered the 'digital age,' an epoch in which the digital—singular—is an entirely new and all-encompassing phenomenon. Rather, its theory, method, and subjects are part of a process of digitalization that is rooted in the discipline's original matters and spreads in manifold ways.

In what follows we will first go back to the origins of the word *digital* and show how it has shaped criminological work from its very beginning. We will then move on to describe the ways in which 'the digital' and its technologies have changed. Criminologists have to take account of these changes when they study specific phenomena, especially because the logics of digitalization tend to tie in with politics and power. Studying digitalization across different criminological fields also warrants a re-thinking of the tools we use for knowledge-making, which we will discuss before we let readers off to browse, delve into specific topics and plan research projects.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-001

Continuities

Maps, tables, sketches, specimen. Criminology has always had an intimate relationship to data and measurement. Like in any other empirical science, infrastructure such as spreadsheets, file cabinets, thermometers, and dictaphones are key to the generation, collection, and organization of criminological data. They are important artifacts in the making of criminological knowledge. Metric tools generate data, but they also act as data when one takes account of the ways in which measurement and counting vary across tools.

Crimes have been counted ever since society "became statistical" between 1820 and 1840, when an "avalanche of printed numbers" swept over many European countries (Hacking, 1990: 1–2), which also manifested itself in emerging governmental rationalities. So when numbers, enumeration, and statistics became part of social science—and governance—'crime' was understood, rendered, and approached in a new fashion (Lomell, 2010). From the beginning, numeric representations of social phenomena were used by social scientists to substantiate theories and to identify statistical laws and regularities, both in order to improve and control 'deviant subpopulations' (Lomell, 2011). Reading regularity and predictability into the first crime rates, Adolphe Quetelet [1796–1874] concluded that the causes of crime must lie outside of each individual's control, and that crime could not be a simple matter of individual choice (Quetelet, [1842] 1996). Crime as a *social* phenomenon—and problem—was invented with the representation of crime in numbers and the creation of the crime rate.

The criminological discourses of the early 1900s were also characterized by debates about the 'correct' instruments and methods for metricizing criminals (Gibson and Rafter, 2006). Today, in the 21st century, debates about measurement tools are not concluded. They may no longer refer to the process of rendering the bodies of criminals into numbers, but to the 'correct' choice of instrument for identifying patterns of crimes and risks. Relatedly, the ways in which criminological knowledge production impacts societies is also not a new discourse. For example, critics have long addressed the discriminatory nature of early metric works such as Lombroso's *Criminal Man* (Basaglia, 1971, cited in Montaldo, 2018). Yet, discussions about discrimination are still central in today's biometric projects (see Biometric Failure by Din and Magnet).

Though it is seductive to think of digital criminology as a new field, the above goes to show that criminologists have a long tradition of engaging (with) digital logics. For many, it may be unusual to think of the above practices of measurement and numbering as digital. The word *digital*, however, goes back to *digitalis*, measuring a finger's breadth, which came to denote discrete entities that can be used to count and calculate with. The discrete units of the abacus are, then, a form of digital information in an *analog* fashion. Yet, today's vernacular refers to digital information as the electronic representation of data in the discrete values of ones and zeros as established in the mid 20th century (see Digital by Wernimont). This conceptual development expresses the continuities that characterize digitalization: that digital technologies and practices

did not arise with the advent of the personal computer or the Internet. They are the result of more comprehensive genealogies.

The way in which *digitalis* relates to concrete, physical units reflects yet another important aspect: the materialities of digital data or digital matters. In the same way in which skulls, maps, and archives are material, electronic data and infrastructures, too, are material. That is to say: a software indicating crime hot spots is not only an extension of analog data, such as maps with needles and pins as prototyped by the Chicago School (Thrasher, [1936] 2013). Its electronic instance, too, is enabled and upheld by a network of humans, items, and material practices without which it cannot exist.

Digital criminology, then, does not refer to virtual and abstract phenomena (Floridi, 2010) or totalizing qualifiers such as 'ages' and 'worlds.' It engages with concrete matters that have specific genealogies. And these material genealogies keep evolving. Infrastructure studies provide a lens to understand these phenomena as evolving sociotechnical assemblages (see Infrastructures by Grisot and Parmiggiani). For instance, rather than studying electronic monitoring in prisons as a discrete technology, one can map the larger and evolving surveillance assemblages and infrastructures that it is part of.

Returning to the word *digitalis*—measuring a finger's breadth—we find yet more aspects that are relevant to defining digital criminology. The activities of measuring and calculation comprehended in *digitalis* have been a part of criminology long before electronic computation arrived. Adolphe Quetelet's correlational approach to crime mentioned above, the anthropometric system of Bertillonage, or Cesare Lombroso's typology of 'born criminals' are examples of measurement, numeric archives, and statistics that were a part of establishing criminology. They are examples of digital methods, of counting, measuring, and calculating with discrete units in the empirical research on crime. Even though biometrics is a word that we mainly associate with today's technologies for facial recognition and border control, Bertillonage is in fact one of the first biometric systems used in law enforcement. Data collection, profiling, pattern recognition as well as the drive for systematization and efficiency are logics that characterized early criminology and still form the basis of electronic digital practices today.

For example, the search for patterns of delinquency in urban environments (Shaw and McKay, 1972), the detection of regularities as performed in routines (Cohen and Felson, 1979) and rational choices (Cornish and Clarke, 1986) are metric and statistical approaches to crime that constitute the logics of predictive policing software today. Some software solutions actually incorporate such theoretical models into their algorithms where they are used to define parameters for the identification of patterns (note that an algorithm is a set of instructions to solve a problem via computation—electronic or not, see Algorithm by Leese). While patrol, too, has always been central to the police, as for instance with Sir Robert Peel's 'Bobby,' the use of advanced statistics led to a prioritization of specific environments over time. The growing amount of digital information available to private and public services as well as electronic computation now enable software-supported predictions (see Prediction by Ķīlis, Gundhus, and Galis). Such predictions are increasingly used to fine-tune police presence and the

prevention of expected crimes, which is captured in the term 'precision policing' (Shults, 2022).

The story of crime patterns, police patrol, and predictions, then, illustrates the ways in which a *digital* criminology involves genealogies, materiality, and calculus. What is more, the turn towards a 'politics of *pre-*' (Kaufmann, 2023), that is the management of security with a focus on the future (Zedner, 2007), was a process that did not arrive with big data and electronic computing, but was also rooted in the history of criminological practices. Think of the role of risk logics in criminal justice that existed long before sentencing was automated, involving large databases (Maurutto and Hannah-Moffat, 2006). Yet, arguably, the role of prediction and prevention became more manifest and integrated into crime control with the advent of the personal computer and digital archives. It is no longer only an *aspiration* to predict and get a sense of control over the future. The availability of predictive instruments suggests a new *responsibility* to predict, prevent, profile, target, and to be efficient. This new responsibility is firmly established in modern societies that are imbued with a preventive gaze, where "incidents are not perceived as tragic accidents, but instead as avoidable events" (Peeters, 2013: 22). The future will stay with us in the future. Prediction will remain a key logic in criminology, because it is "malleable enough to be adopted and adapted by several criminal justice actors, yet robust enough to maintain a common identity and concept across different organisational sites and applications," as Ķīlis and colleagues elegantly put it in their contribution (see Prediction by Ķīlis, Gundhus, and Galis, cf. Star and Griesemer, 1989).

Creating profiles of criminals, terrorists, or 'dangerousness' is another phenomenon that tends to be presented as an innovative, 21st-century security practice. But here, too, historical developments and continuities characterize today's trends. The practice of measuring and categorizing criminals, victims, and human behavior started in the early 1900s. Early biometric instruments set the trend to measure anything from head shape, weight, body build and to correlate it with behavior in order to arrive at a profile of criminals by type of crime. While modern profiling rarely seeks to predict behavior, the idea of using categories and profiles in order to control crime is as present as ever. Profiles are used, for example, to flag suspicious financial transactions, conspicuous browsing behavior, or to predict the skin color of a suspect from DNA (see DNA/Big Genome Data by Kaufmann).

For criminologists it is thus important to remember that digital technologies have a history and that criminological researchers are, to a certain extent, well-equipped to study digital technologies and practices. But it would be incorrect to claim that digitalization is a development without interruptions, jumps, and changes. As Roger Burrows puts it: "the 'stuff' that makes up the social and urban fabric has changed" (2009: 451).

Changes

If the 'stuff' that makes up digitalization has changed, the advent of electric power arguably introduced one of the most significant shifts. With a simple abacus, for example, one would be able to calculate multiplications of 10-digit numbers in one second. The British *Colossus*, the first programmable, electronic, digital computer, was used in World War II for cryptanalysis. By computing 5,000 operations per second, it was able to decrypt the German Lorenz cipher and changed the course of war. Since 2022 machines exist that are capable of calculating 10 to the power of 18 operations per second. Such supercomputers are in need of energy comparable to that of 16,000 homes (Johnson-Groh, 2023). Today's supercomputers, for example, simulate anything from the universe to nuclear weapons capabilities which form the basis for security politics of global dimensions and beyond. We can observe that electric power, powerful computing, and powerful political dynamics change together, something that is crucial to criminological analysis. But how do these shifts connect?

First, we can mention the phenomenon of *datafication* (see Datafication by Chan). The above examples show that ever-more aspects of life become captured as information, or to put it differently: informationalized or datafied. With the spread of digital infrastructures more objects, bodies and behavior become subject to measurement. More measurement activity generates more data, and vice versa. Datasets can be distinctive, specific, and small. But they can also be various, velocious, and voluminous. While we have not left filing cabinets and paper folders behind, electronic datasets have become so massive that their storage is a considerable percentage of global carbon emissions. Though their size is no longer graspable, which is why we like to call it 'the cloud,' we can still observe that data are material and create very concrete effects. 'Big data'-centers are so physical that they change the atmosphere and the 'stuff' needed for electronic technologies has led to a competition over metals, mining, and markets.

Unlike 'the cloud' suggests, electronic data are not ever-present, they are not everywhere. Some societies are heavily digitalized, while others are not. To some populations digital technologies are the foundation of their social lives, while others are excluded, underequipped, or disinterested. Nonetheless, digital data and technologies have grown in relevance to the extent that most people experience the effects of datafication. This has brought about technological innovation, but has also led to new crimes and harms, exponential surveillance and sousveillance.

A familiar example would be social media, which have become integral to many peoples' lives. Social media have reshaped social relations and politics, and thereby also crime and control (see Social Media by Twigt). They allow users to engage in harmful or illegal acts, such as online mobbing, grooming, or drug trade. Social media provide platforms for new social practices (e.g., sex work) and groups (e.g., incels). But they also generate new dynamics, such as political action, radicalization, control practices, and the resistance to control. Another example is the 'Internet of Things' (IoT),

which refers to the ability of everyday objects to connect to the internet, other communication networks, and each other to create a 'smart' environment (that is Google's vision mentioned above). These objects generate and exchange data via sensors and act based on data analysis. Many IoT devices, however, have not been designed with privacy and social implications in mind; they follow a technical, not a socio-technical approach to technology development. Needless to say that the IoT is per se a surveillance tool, which can be abused for both control and criminal purposes (see Internet of Things by Milivojevic). The IoT is also the base model for 'smart cities'—technologically sophisticated urban areas that use digital devices, data analytics, and innovative architectural design to overcome traditional metropolitan challenges such as traffic congestion, crime, and incivility. The ways in which such projects root entirely new surveillance infrastructures in societies, facilitating new forms of crime management, but also social control and types of crime, is a phenomenon of rising importance to digital criminologists (see Smart City by Hayward). In tandem with a rise of smart infrastructures we see a normalization and dispersion of digital borders that co-shape the politics of inclusion and exclusion, wanted and unwanted mobilities, and knowledge-making about moving subjects and dynamic territories (see Borders and Border Control by Jeandesboz).

Indeed, studying digitalization also means researching how data and technologies impact the *temporal and spatial dimensions* of criminological phenomena. Speed and heightened connectivity do not always imply sophistication or betterment. Rather, they redefine criminological fields of study in various ways. For example, marketplaces for drug trade still involve farms, cityscapes, and transport networks, but also online platforms, e-payment systems, and the 'darknet.' The darknet is a term adopted by law enforcement to describe those parts of the internet that are encrypted and layered to enhance privacy. With increasing digitalization drug markets may not necessarily become more global (in fact, many of them were dependent on global connections long before the rise of online drug markets), but digitalization leads to new forms of both closeness and distance. Sellers and buyers get in touch without having to leave their homes and the darknet changes the security of buyer–seller contact for better and worse. In addition, strategies for advertisement, negotiation, and feedback are adapted to digital markets. This example illustrates the complexities that electronic data and technologies bring to criminological phenomena.

To grasp the relationship between digitalization, crime, and crime control, then, it is important to understand the properties of digital devices and what they afford (Gibson, 1979). An affordance is that which enables and disables, encourages or discourages specific actions. Environments, objects, or artefacts do not cause behavior per se; rather, they offer, enable, or *afford* some actions over others. What technologies afford is a key entry point in understanding what humans can, and will, do with them (see Affordances by Wood and Arpke-Wales). Affordances are specific to each object. In digital criminology, entire research projects can be dedicated to studying the affordances of digital devices, how they change environments, practices, and other things. Such de-

vices can include computers and software, or credit cards and banking infrastructure, just to name a few.

In order to study the effects of digitalization more generally, it is helpful to understand the affordances of electronic digital data (Kaufmann and Jeandesboz, 2017). Above, we have pointed out that digital data are *discrete* (as opposed to continuous), which makes it easy to count, calculate, and measure with digital data. Hence, the digital character of electronic information is the basis for numeric and computing operations that are needed to run anything from local offline data programs to complex telecommunication platforms (see Computation by Mazzilli Daechsel). What is more, digital information can be *stored*, which means that we have archives of entirely new dimensions. Paper archives were the big data of the early 20th century. Today's big data and digital archives, too, are physical as they come in the form of hard disks and data centers. Actually, some of today's big datasets also involve paper forms that are translated into electronic data and across professional domains (see Translation by Wilson-Kovacs). The digitalization of archives, however, alters the answers to questions such as: Under which conditions are digital archives gathered and made available? Who governs them? And how do they reshape ontologies and epistemologies of crime and crime control? (see Archives by Thylstrup and Veel).

Digital archives make information *available* and *searchable* to different parties, which also means that it can be used for simple and complex analyses. Due to its electronic form, data are also *transferable* if systems and rules are set up for this transfer. The speed with which digital data can be transferred is, for example, a crucial ingredient to online communication. Digital electronic data afford the building of *networks*, which means that new group formations can occur. They are also *traceable*, which is key not only to communication and networking, but also the reason why any open data transfer can be surveilled, making it possible to capture and store information. Crucially, digital information is made, unmade, and *co-created* by people and machines in different environments (Kaufmann and Jeandesboz, 2017). This means that digital criminology is not the study of society *or* technology, but of socio-technical phenomena. Technologies do not emerge without context. Designing, building, using, even destroying technologies is always a process that involves humans and technologies alike. All technologies—an encrypted network, a platform for policing or sex work, a messaging app, a database or an electronic shackle—have design decisions built into them that are part of shaping how digital practices and dynamics unfold. Humans and technologies are entangled in that they shape one another. Each of them has agency (see Agency by Krasmann). Technologies and their use, then, need to be understood in their *specificity*, that means in the specific contexts in which they appear and take effect.

Crime, victimization and digitalization

Paying attention to agencies and affordances can help us find entry points for studying digital aspects of crime and victimization. Digitalization affects most fields of criminol-

ogy (and if it does not, its absence may also be an interesting aspect to study). Consider how online markets and banking change financial crime and surveillance (see Financial Crime and Surveillance by Amicelle). Electronic gadgets enable new forms of identification, but also ID-theft (see Identity theft by Langford, Svensson, and Wærstad). Synthetic data can be used for enhancing privacy, but also for mis/information practices or the generation of illegal contents (see Synthetic Data and Generative Machine Learning by de Vries). Social media can enable social networks that have a preventive effect on their members, but also propaganda, radicalization and 'networked hate' (see Hate Crime and Networked Hate by Cameron, Stratton, and Powell), violence, vulnerability, or victimization.

Studying digital dimensions of crime is not the same as studying cybercrime. Cybercrime is often used as an umbrella term which includes all offenses either occurring in or being facilitated by an online environment. Research on cybercrime is very established in criminology. Yet, its focus on online phenomena implies that offline offenses fall out of its scope (Cybercrime by Holt and Holt). Digital criminology, as outlined in this handbook, underlines that a study of online crime would take offline, analog, or other aspects that enable criminal practice into account. What is more, a strict definition of cybercrime would relate to crime only. However, since digitalization is a process of constant change, definitions as to what counts as crime, what is criminalized and in need of regulation are under constant development, too. When using 'digitalization' instead of 'cybercrime' as a vantage point we can take account of these dynamics and ambiguities.

New forms of online vulnerabilities (see Vulnerability by Ranchordas and Beck) emerge hand in hand with new forms of online victimization (see Victimization by Walklate). While not all examples of victimization are an offense, they may become an offense in need of legal regulation. The same is true for online abuse. Not only do practices such as sexting or revenge pornography challenge traditional cultural conceptions of abuse especially when it comes to age, gender, and the status of the 'deserving' victim (see Abuse by McAlinden). But much of such online behavior still has to be understood and established as criminal offense before it is integrated into law. That is to say, digitalization enables versions or varieties of crimes that first have to be identified as those. For example, as more and more administrative fields are digitalized, new forms of ID-theft occur. The above underlines the importance to study how digitalization alters criminalization processes.

Even when a certain digital practice is established as an offense, it is not necessarily clear who will be treated or targeted as the offender: is the individual or are entire platforms and their owners liable? It is thus crucial for criminologists to take into account how digital and online devices influence classic criminological themes of criminalization and liability. At what point is the sharing of sexualized content a form of abuse, who is the offender and what is the role of digital technologies? When is hacking a creative and progressive practice and at what point does it turn into an offense (see Hacking by Wall)? How does the migration of drug markets to the darknet create both positive and negative effects on drug consumption? How can law-making take ac-

count of the two-sidedness of the darknet as it provides encryption and safety from surveillance on the one hand, and a haven for criminal activity on the other (see Darknet by Tzanetakis)? How are control and care entangled in the design of robots, and (when) can robots be considered victims of crimes and violence (see Robots by Lintvedt and Sandvik)?

While many of these issues have been present in law-making for a long time, digital technologies create new practices and dilemmas, because they are characterized by a different set of affordances that allow for a variety of usages. Online digilantism, for example, is the use of the internet to act on perceived offenders without legal authority (see Digilantism by Trottier). An example are social media groups that take initiative to report what they consider as offensive or confront the alleged offender themselves. Digilantism is different from vigilantism, its offline counterpart, because it is mediated by networked technologies that allow their campaigns to 'go viral.' When is digilantism, then, a positive form of social online engagement and at what point does it turn into networked hate with a potential for offline confrontation and violence? And how does the digital environment enable or disable this?

Using digitalization as a vantage point for empirical studies is thus also a way of acknowledging the complexities of phenomena. One example is sex work, where the internet, smart phones, and cameras play an increasingly central role. By going online, a new market for sex work has been created and new powerful actors, for example platform owners, emerge (see Platforms by Egbert). These can act as both protectors or gatekeepers for sex workers. Digitalization generates new possibilities for sex workers, such as audiences and fans, but also vulnerabilities such as being recorded without consent, or they face risks of being exploited by platform owners (see Sex Work by Rand).

Crime control and digitalization

Electronic digital data afford activities of measuring, calculation, and tracing, which means that they also afford new means of crime control. All three are key to the logic of surveillance, of collecting information about individuals, groups, and behavior, and of categorizing this information into risks and predictions. While surveillance has always existed in analog and non-digital forms (see Surveillance by Lyon), it has recently been joined by concepts of dataveillance (Clarke, 1988), capture (Agre, 1994), surveillant assemblages, and data doubles (Haggerty and Ericson, 2000). All of these concepts can be questioned and developed further. What their emergence indicates, however, is that the exponential rise of electronic data has led to an exponential rise of surveillance. Such growing volumes of anything from metadata to free text (e. g., in messages) are in need of evermore sophisticated data processing programs. Algorithms can be trained with more, or less, human interaction, the latter of which we find in machine learning or Artificial Intelligence (AI) models. AI is not new, but evermore powerful computers can analyze large amounts of both structured data (i. e., data organized in

categories) and unstructured data (un- or less organized data). All of these developments have also increased the use of AI in crime control (see Artificial Intelligence by Van Brakel).

Police intelligence, for example, has a long history, but technological advancements are here of particular relevance due to the central role that information plays in intelligence practices. New trends and surveillance solutions enhance collecting, storing, analyzing, and presenting data and information (see Intelligence by Gundhus and Lundgaard). These developments also link intelligence analysis to other digital technologies, such as prediction software and data integration platforms that are adopted across police institutions, including patrol. Symptomatic of this shift is the rise of strategies such as predictive policing and precision policing, as mentioned above. Overall, the volumes and combinations of datasets, advances in predictive analytics and data mining, as well as competitive markets in security technology, have all congealed to mobilize an intensive datafication of police work that continues to accelerate (see Policing by Wilson). The growth, impact, and meaning of facial recognition technology in policing is another example that represents a dramatic shift in police surveillance capability (see Facial Recognition by Fussey). Here, public institutions are increasingly dependent on private technology providers whose range of products tend to cover much more than law enforcement applications. As a consequence, new meeting points between different public and private surveillance practices emerge.

The digitalization of crime control extends to jurisprudence. Courts used to be synonymous with central, imposing buildings where proceedings took place in the presence of the parties and before a judge. Today, courts are moving online and employ AI for decision-making (see Online Courts by Mentovich and Einy). While these changes may render proceedings more efficient, they are also met with issues of representation and unequal access to digital infrastructures. Sentencing algorithms, too, may provide efficiency and structure, but also perform hidden and obvious forms of discrimination (see Sentencing and Risk Assessment Algorithms by Ugwudike). A key issue is thus that digitalization entrenches unjust systems and discriminatory practices in legal domains, which introduces new power dynamics. These dynamics call for new concepts of justice, as for example captured in the term *data justice* (see Data Justice by Redden)

Punishment, too, is being digitalized, where surveillance plays a key role. An example that illustrates the transformative power of electronic surveillance may be the footcuff, originally an iron ball connected to a prisoner's foot, used to prevent escape. The footcuff is something that assists in, but is not per se surveillance. When it comes to prison surveillance, Jeremy Bentham became famous for his panopticon. Its architecture allows guards to watch inmates at any point in time without the inmates knowing when the guard is watching. Today, the footcuff and the panopticon are combined and built into ankle bracelets that carry sensors to remote-trace convicts (see Punishment by Lunde and Scharff Smith). The advantages, challenges but also promises and desires related to digital prisons will have to be documented and gauged in ongoing assessments.

Indeed, the digitalization of crime control needs to be understood in terms of its expectations and consequences. And the expectations that law enforcement has towards *big data* are big (see Big Data by Završnik). Terms such as the 'data deluge' (Bevan, 2015) and 'data tsunamis' (Rubinstein, 2013) suggest that data are liquid and can take any form. The shape or the patterns that (we make) data reveal form the basis for taking action. For that reason, patterns are adopted all across law enforcement. As illustrated above, patterns influence policing (where to focus police attention), intelligence (identifying suspicious patterns), sentencing (identifying those likely to reoffend)—and much more.

Association is here a key analytic practice, which is also used to process analog information (cf. Kaufmann, 2023). To associate is to join, to make a connection "in an interest, object, employment or purpose" (Harper, n.d.), which also could be done on paper or with simple calculations. With the rise of digital information, however, association has shifted in terms of reach, quality, and scale. Very large datasets can no longer be managed and analyzed by humans alone, which is why computation and algorithms have become so important. Algorithms associate different datasets with each other (see Kaufmann, 2023). The most common approach of associating different datasets with each other follows a Boolean logic (Kitchin, 2016), named after the mathematician George Boole. We know them as if-then rules, that is: when *if* is true, *then* is executed. Here, patterns are identified via correlative logics. Another type of association is to generate patterns as we find in so-called 'self-learning' algorithms and large language models like Chat GPT. Association has become a central aspect of surveillance. It is critical to any kind of categorization, sorting, and profiling that we experience on an everyday basis in public management, including law enforcement, health policy, welfare schemes, migration, and border control. All of them are key areas of societal organization, where categories and patterns are expected to produce actionable knowledge (see Categorization and Sorting by Franko).

The implications of the molecularization of society, as Nikolas Rose observes them, are also reflected in digitalization, producing a shift in law enforcement: away from a *deep ontology* aimed at studying causalities, towards a *flat ontology* concerned with circulations and correlations (Rose, 2000). That is to say, thinking of society at a molecular and digitalized level places the focus on management, correlations, and surfaces rather than depth and causalities. The rising confidence in patterns as a base for managing societal relations is expressive of this flat ontology. The propensity of digital data to aggregate, the hunger for evermore-complete datasets and their analysis favor patterns, profiles, and categories over individuality. The growth of databases (see Databases by Bellanova) and the need for automation to analyze them (see Automation by Mann), too, are symptomatic of flat ontologies. They replace the attempt to capture "a set of circumstances in the world" (Amoore, 2011: 32) with the logic of the *derivative:* "the data derivative is exposed to the underlying data without collecting them, created across the gaps and absences, in the interstitial spaces of inference and expansion" (Amoore, 2011: 33). The derivative does not reflect 'circumstances in the world', but it is an analytic device to manage, amongst other things, offenders, crime, and pathology.

One ambition is here to analyze large datasets in order to enhance precision and to arrive at the particular. Today, this even involves the generation of synthetic data in order to increase accuracy (see Synthetic Data and Generative Machine Learning by de Vries). However, the particular slips away in the process of creating actionable knowledge in the form of a data derivative, a pattern, a type (Hopman, 2023). Hence, increasing differentiation and larger datasets do not lead us to the particular, but only to its increasingly differentiated type. How such derivatives, patterns, and types circulate in specific criminological domains, how they change crime and crime control is yet another central aspect of digital criminology. These rationalities are entrenched with a range of challenges and power-dynamics, which underlines the many ways in which digitalization is political.

The politics of digitalization

Digitalization requires investments. *Budgetary policy* prioritizes certain forms of digitalization over others. This is a field that becomes relevant to criminologists when they study law enforcement institutions, where tech adoption, institutional culture, and budget decisions define digitalization processes. The costs of digitalization also affect related domains through knock-on effects. Labor markets and work cultures are classically impacted by the development of technologies. Here, the thrive for efficiency and effectiveness is also challenged by worries about a compromise of competence and dispensation amongst personnel, something that criminologists have also studied in relation to the digitalization of police work or court rooms (Ratcliffe et al., 2020).

But there are also more overarching *economic dimensions* to digitalization. Most digital infrastructures are owned by private companies and public–private partnerships have long histories in law enforcement. And yet, the capitalist role of big tech companies and related market economies increasingly define the ways in which technologies are used in law enforcement. Whether in punishment and sentencing, police and forensic work—providers of technological solutions are normally private. Commercial interests of selling and developing products render both private users and public institutions into customers, where knowledge-making and expertise is co-produced by public and private actors, if it is not outsourced entirely (see Privatization by Lomell). When law enforcement becomes increasingly digitalized, it is indispensable that criminologists take account of the role that private companies play in prosecution, surveillance, and data work in organizations. While some digital solutions do not use customer data for training and development, many private technologies are data-hungry. Not only does that turn user data into the prime capital, but it also places surveillance at the base of technologies' functioning (Zuboff, 2019).

In many ways, however, data have also become less available or traceable, which is where *black boxes* or *blackboxing* comes into play (Latour, 1999; Pasquale, 2015). The changing relationship between the public and the private is a central aspect here, too. Due to the collaborations of public and private actors, public data are not neces-

sarily open data in the sense that they are publicly available. In the history of information much data owned by public institutions was accessible to specific public entities only, e.g., the police. However, with the rise of electronic infrastructures, access becomes a key issue: a lot of information is owned and managed by private companies that enable and restrict data access.

For individuals it has become impossible to own or even access electronic data about themselves. One reason for that is the complexity of digitalization as digital traces of individuals spread across different services and devices. Another one is that data are integrated into bulk analyses, traded, re-used, and appropriated as capital by providers of services which makes it impossible for individuals to own or even trace information. What is more, accessing electronic data requires know-how and infrastructure, which becomes a challenge for individuals, public institutions, and even companies depending on the type of data to be accessed. Here, new vulnerabilities and vulnerable populations emerge. A second issue that complicates access is the changing capacities and practices of calculation. Data analysis has become complex, surpassing human ability to follow. What is more, in many cases access to datasets and knowledge about calculative processes are business secrets, a black box by choice and a result of privatization. These dynamics of in/visibility impact discretion: decisions are co-shaped by algorithmic logics and (semi-)privatized institutions, something that is particularly sensitive in the field of security, crime control, and law enforcement.

Due to their mathematic form, it is seductive to think of computers and algorithms as neutral instruments (Goffey, 2008: 16). However, the initial design and training of algorithms are the result of socio-technical interactions, which necessarily creates machinic *bias and discrimination*, whether positive or negative (see Bias by Oswald and Paul). Bias, then, is everywhere. Even with the increase in data and computation technologies, data analysis has not become more neutral. Current debates about the type of databases used or the parameters of algorithms illustrate that computation is and will be a socio-technical issue. Claudia Aradau offers the analytic concept of *error* as a force for development. It was the errors in the early (anthropo)metric projects that led criminologists to continue to engage with digital systems, including today's electronic solutions (see Error by Aradau). Error is also related to failure, a term that Shoshana Magnet and Ravida Din use to describe the negative bias and the discriminatory nature of today's surveillance systems (see Biometric Failure by Din and Magnet). This can include failures to identify specific skin colors, the exclusion of people with disabilities from digitalization, forms of categorization that are racist and produce disproportional or unfair effects, as well as algorithms that do not pick up on certain forms of victimization. Data have become more than big, however, in tandem with the challenges of skewed data collection. Here, the discrete nature of digital data also contributes to the issue of discrimination. Digital, as mentioned above, refers to measurement in discrete units. In its electronic format, too, digital data are expressed in ones and zeros. The inability to capture anything on a continuum—or to put it differently: the need for

categorization—necessarily leads to types, profiles, and groups and the risk of creating discriminatory categorizations is high.

Catch-phrases such as FATE, which stands for Fairness, Accountability, Transparency, and Ethics (Shin and Park, 2019), are applauded vantage points for the *regulation of big tech*, but notoriously hard to implement. The issue of regulation brings us back to the point that large databases and proprietary software are key to many private companies' business models, protecting businesses from having to release their algorithms or publicizing their databases. This complicates legal regulation. In fact, regulation tends to be privatized as terms and conditions are written by tech owners. Companies effectively become the regulators of digital information flows and digital behavior, increasingly appropriating domains originally governed by public law and criminal justice agents. Lawmakers are then left with limited options for contents to regulate. Effectively regulating such contents would also require streamlined international laws in order to be effective. Such laws would also require all companies to enact transparency in a way that actually enables an understanding and the regulation of technologies. However, the fact that solutions are increasingly complex and complicated to understand—even for their owners—does not make the regulation of technologies any easier. The European AI Act (European Commission, 2021), for example, places the main emphasis not on regulating technology development per se, but on risks related to technology usage. These risk categories would range from minimal to unacceptable risks, where the "biometric identification and categorisation of people" (European Parliament, 2023), for example, would count as unacceptable. However, not only is this risk category so broad that many technologies already in use would be considered unacceptable, but a prohibition of their use would also hamper market competition. While both of these effects can be positive, it remains to be seen how regulation can and will be implemented in practice and whether Europe is able to attenuate the powerful influence of Silicon Valley and equally powerful tech hubs in Africa and Asia.

Regulating the use of technology for surveillance purposes, meaning limiting *privacy intrusion* and defining the threshold to suitable, necessary, and proportional use of surveillance, is yet another issue that has been subject to a long-standing debate. The General Data Protection Regulation (GDPR) implemented by the EU is here often cited as the highest standard privacy regulation in the world (European Parliament, 2016). While it does prevent some companies from operating in Europe, the use of surveillance for security purposes is the argument that creates a loophole in any privacy regulation. Any of these regulatory aspects are important to navigate and study when criminologists assess the use of technology in their field of expertise (see Privacy and Data Protection by Bygrave).

Studying digital phenomena

Studying digital technologies, practices, and environments does not only require analytic concepts, but also suitable methods. Generating insight into digital practices

also invites or requires researchers to navigate technologies as research tools. Surveillance studies and digital sociology have here paved the way for adopting new methods and reflecting about the ways in which digital technologies are part of making criminological knowledge. Deborah Lupton, for example, traced how many traditional methods in sociology have become digitalized, such as interviews, surveys, or observation (2014). When studying documents and other types of texts, Rogers distinguishes between 'digitised data objects,' those that existed as film, photo, audio, text etc. before digitalization, and those that are 'native' to the electronic digital contexts (2014). Though intuitively one would place the analytic focus on databases, software, or AI-generated contents to understand digitalization, there is value, as this chapter argues, to understand which non-digital, low-tech, analog, or offline technologies are part of such 'native' objects (see Low-Tech by Vestad). At the same time, it is of analytic importance to take account of the specific properties of electronic digital objects to understand how and why they work in the making of criminological knowledge. Here, the fields of netnography (focusing on online environments) and digital ethnography (including any digital tool) open different doors to understand digital cultures and phenomena (see Online Ethnography by Gibbs and Hall). Digital criminology, then, is not just a study of humans that use technologies, but also a study of digital devices and online phenomena: how do we make a digital object speak? Adams and Thompson suggest gathering anecdotes about digital objects, following actors, understanding affordances or 'invitational qualities' of digital technologies, studying breakdowns, translations, and passages, that is studying how technologies become what they are (see Interviews with Digital Objects by Adams and Thompson). In such contexts, the laboratory is also an important but largely understudied site in criminology. Labs play an important role in crime and crime control, but also in the development of the discipline and its scholars. Laboratories do not just produce new technologies, practices, and knowledge. They also produce 'new types of people,' that is subjectivities expressive of ongoing societal and professional digitalization (Schmidgen, 2021; see Labs by Mazzilli Daechsel).

Understanding how digital objects are part of creating criminological knowledge is also key when choosing a specific digital technology for establishing access to a group, or for conducting interviews. For example, while one could make offline, face-to-face interviews with drug consumers, the relative anonymity of encrypted sites can be an alternative that allows for different interview dynamics, which can lead to different or new insights. These interviews can be held on gaming software and other encrypted channels, or texting apps—just to name a few (see App-Based Textual Interviews by Bakken). Sex work, too, may be understood differently with different methods. While research on sex work has a long and important history of studying offline environments, sex work itself has also moved online. This means that researchers should also pay attention to chats and forums for understanding online groups and practices of within their specific field (see Researching Online Forums by Šupa). Indeed, seeking access to relevant fora or social media is crucial when conducting research on phenomena that are regarded as digitally native, such as online radicalization. When doing so, navigating the balance of public and private space, or even using one's own accounts

for such research projects requires careful planning (see Recruitment via Social Media by Andersen). An alternative way of approaching online radicalization could also be from the side of those who seek to regulate it, their online presence, and their use of tools for the detection and analysis of user patterns.

A tool that is suitable for both qualitative and quantitative projects is the net crawler. They are particularly useful if the aim is to collect a lot of structured information from the internet, that is content data presented in the same patterns, i. e., a chatroom with date, time, a line for a heading, and a content box. A crawler is a program that can be used to automatically collect or 'scrape' this content. Often, however, these are large datasets that require structured storage and analysis with yet different software solutions. Not only is it crucial to be aware of the variations different software solutions can produce in analyses, but it is also key to reflect about the ethics and the invasiveness of crawling data that is not produced for the purpose of doing research. As with researching online forums, such research requires careful reflection about the relative privacy and personal character of information shared online despite it being publicly available. Alternatively, it is also an option to look at services that have already conducted statistical analyses, such as Google Trends or n-grams, for example for checking the use of specific terms over time. It is, however, important to reflect about the limits of this pre-given information and the ethics of using big tech services for doing research. There are many ways of using automation and digital tools for doing research. Some designs, for example, also implement experiments that expose people to robots. Such research designs require careful and ongoing ethics monitoring (see Ethics by Markham).

There are plenty of ways of adopting digital objects and methods into criminological research: by using simple digital devices to record conversations, by coding pictures or videos, by analyzing algorithms, by studying or building databases, by using social media actively, by using the encrypted net or 'darknet' as a site for research, by following discussion forums, capturing websites, collecting tweets/Xs, or following digital devices around in everyday practices. Art, too, is an entry point to knowledge production and reflections. Here, art projects can embrace speculation more freely than classic research methods. Through this, art can create new insights, but also prompt commentary and critique (see Art as Method by Dewey-Hagborg).

What is key to all the above-mentioned methods and digital devices for knowledge production is to consider their ethical implications: Why should it be fair to use data for research that are produced by others in a non-academic context? Is my approach extractivist, i. e., do I harvest or mine data in an unfair or destructive fashion? Is the method I choose safe for everyone participating in research, including myself? How do I protect data and when do I cross the line into private spheres? Do the tools I use for analysis do the phenomenon I study justice? In many ways, these questions apply to analog and digital methods alike. But it is their 'digitality' that needs to be a point of reflection, that is: How do the properties of electronic digital data and environments change methods, practicalities, and ethics? A thorough understanding of the technologies we use to do research is key to reflect about the ways in which

they shape access, participation, interactions, communication patterns, research data. Good knowledge of one's digital research tool is key to assess their advantages and limits—and most important—reflect about one's own co-creative role in the research process (see Accessing Online Communities by Kaufmann). Technologies are not only part of our attempts to understand social practices and phenomena, but they also shape the knowledge we create about them.

Openings

The ambition of digital criminology is to establish itself "within the discipline, not as a sub-discipline" (Powell et al., 2018: 190). Digitalization is a phenomenon that cuts across all criminological fields, affecting criminalization, crime, and crime control alike. Digitalization is not a new phenomenon. The original meaning of the word *digital* relates to measuring, counting, and calculation. These were equally relevant for the metric and statistical ambitions of early 20th-century criminology as they are today when complex computing technologies continue to re-shape societies. Hence, we invite readers to examine the continuities and changes, as well as the materialities and socio-technical activities that characterize digitalization. We invite readers to help characterizing 'the digital' further by studying processes of digitalization in their specific contexts, moving beyond simple online/offline, analog/electronic, high/low-tech distinctions. Questioning linear accounts of digitalization, ambiguating categorizations, and transgressing epistemological borders is what digital criminology can contribute to the study of technology, crime, and crime control (see Borders and Border Control by Jeandesboz).

This book should be read as a real handbook, a guide for crafting knowledge, in that it provides a glossary with concepts, tools, themes, and methods that equip readers with the key definitions of a term and suitable further readings. The book is edited to inspire. It encourages readers to combine analytic devices with methods and themes that are relevant to their area of interest. While it could seem like some entries relate to either crime or crime control, method or theory, many chapters touch upon several cross-cutting aspects, which is why this handbook's chapters or terms are presented in an alphabetic list rather than in sections. We see value in finding the ambiguities in both the 'digital' and the 'crime' in 'digital criminology' as they vary across different fields and jurisdictions.

Since digital criminology relates to fast-moving, diverse, and dynamic phenomena, handbooks are necessarily incomplete; neither do all authors fully adopt each other's analyses. We thus expect readers to question terms, develop them further and add new ones. What is more, we also want to push readers to consider how their own findings can challenge and change the politics and dynamics of digitalization. This can be done by complicating easy distinctions between crime and crime control, or by providing pathways for tackling discrimination, exploitation, and surveillance creeps, for avoiding victimization, reducing vulnerabilities and injustices that an emerge in tandem with digitalization.

References

Agre, P. E. (1994). Surveillance and capture: Two models of privacy. *The Information Society*, 10(2), 101 – 127.

Amoore, L. (2011). Data derivatives: On the emergence of a security risk calculus for our times. *Theory, Culture & Society*, 28(6), 24 – 43. https://doi.org/10.1177/0263276411417430.

Basaglia, F., & Ongaro Basaglia, F. (1971). L*a maggioranza deviante.* Turin: Einaudi.

Bevan, A. (2015). The data deluge. *Antiquity*, 89(348), 1473 – 1484. doi:10.15184/aqy.2015.102

Burrows, R. (2009). Afterword: Urban informatics and social ontology. In M. Foth, *Handbook of Research on Urban Informatics: The Practice and Promise of the Realtime City* (pp. 450 – 454). Hershey: IGI Global.

Clarke, R. (1988) Information technology and dataveillance. *Communications of the ACM*, 31(5), 498 – 512.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4), 588 – 608.

Cornish, D. B., & Clarke, R. V. (1986). *The Reasoning Criminal: Rational Choice Perspectives on Offending.* New York: Springer-Verlag.

European Commission. (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (Accessed 16 February 2024)

European Parliament. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04 (Accessed 16 February 2024)

European Parliament. (2023). EU AI Act: First regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (Accessed 16 February 2024)

Floridi, L. (2010). *Information. A Very Short Introduction.* Oxford: Oxford University Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Gibson, M., & Rafter, N. (2006). Editors' introduction. In C. Lombroso (ed.), *Criminal Man* (pp. 1 – 36). Durham, NC: Duke University Press.

Goffey, A. (2008). Algorithm. In M. Fuller (ed.), *Software Studies – A Lexicon* (pp. 15 – 20). Cambridge, MA: MIT Press.

Hacking, I. (1990). *The Taming of Chance.* Cambridge: Cambridge University Press.

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology*, 51(4), 605 – 622.

Harper, D. (n.d.). Associate. Online Etymology Dictionary, https://www.etymonline.com/word/associate#etymonline_v_17972 (Accessed 24 November 2022)

Hopman, R. (2023). The face as folded object: Race and the problems with 'progress' in forensic DNA phenotyping. *Social Studies of Science*, 53(6), 869 – 890. https://doi.org/10.1177/03063127211035562.

Johnson-Groh, M. (2023). *Achieving More with Less: Optimizing Efficiency in Supercomputing.* Simons Foundation. https://www.simonsfoundation.org/2023/02/22/achieving-more-with-less-optimizing-efficiency-in-supercomputing/ (Accessed 16 February 2024)

Kaufmann, M. (2023). *Making Information Matter. Understanding Surveillance and Making a Difference.* Bristol: Bristol University Press.

Kaufmann, M., & Jeandesboz, J. (2017). Politics and 'the digital': From singularity to specificity. *European Journal of Social Theory*, 20(3), 309 – 328.

Kitchin, R. (2016). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14 – 29.

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies.* Cambridge, MA: Harvard University Press.

Lomell, H. M. (2010). The politics of numbers: Crime statistics as a source of knowledge and a tool of governance. In S. G. Shoham, P. Knepper, & M. Kett (eds.), *International Handbook of Criminology* (pp. 117 – 152). Boca Raton: CRC Press.

Lomell, H. M. (2011). Making sense of numbers: The presentation of crime statistics in the Oslo police annual reports 1950 – 2008. In A. R. Sætnan, H. M. Lomell, & S. Hammer (eds.), *The Mutual Construction of Statistics and Society* (pp. 191 – 206). New York: Routledge.

Lupton, D. (2014). *Digital Sociology.* London: Routledge.

Maurutto, P., & Hannah-Moffat, K. (2006). Assembling risk and the restructuring of penal control. *The British Journal of Criminology*, 46(3), 438 – 454. https://doi.org/10.1093/bjc/azi073

Montaldo, S. (2018). Lombroso: The myth, the history. *Crime, History and Societies*, 22(2), 31 – 61.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information.* Cambridge, MA and London: Harvard University Press.

Peeters, R. (2013). *The Preventive Gaze. How Prevention Transforms Our Understanding of the State.* The Hague: Eleven International Publishing.

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology.* Abingdon: Routledge.

Quetelet, A. ([1842] 1996). On the development of the propensity to crime. In E. McLaughlin & J. Muncie (eds.), *Criminological Perspectives. Essential Readings* (pp. 23 – 39). London: SAGE Publications.

Ratcliffe, J., Taylor, R. B., & Fisher, R. (2020). Conflicts and congruencies between predictive policing and the patrol officer's craft. *Policing and Society*, 30(6), 639 – 655. DOI: 10.1080/10439463.2019.1577844

Rogers, R. (2014). *Digital Methods.* Cambridge, MA: The MIT Press.

Rose, N. (2000). The biology of culpability: Pathological identity and crime control in a biological culture. *Theoretical Criminology*, 4(1), 5 – 34.

Rubinstein, I. (2013). Big Data: The end of privacy or a new beginning? *International Data Privacy Law*, 3(2), 74 – 87. https://doi.org/10.1093/idpl/ips036

Schmidgen, H. (2021). The laboratory. *Encyclopedia of the History of Science.* doi: 10.34758/sz06-t975

Shaw, C. R., & McKay, H. D. (1972). *Juvenile Delinquency and Urban Areas.* Chicago: University of Chicago Press.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277 – 284.

Shults, Chief J. F. (2022). *Precision policing: The next law enforcement era.* https://www.police1.com/investigations/articles/precision-policing-the-next-law-enforcement-era-q4cwlki2hY9DyJTY/ (Accessed 16 February 2024)

Smith, D. (2015) *Google Chairman: 'The Internet Will Disappear'.* Business Insider. https://www.businessinsider.com/google-chief-eric-schmidt-the-internet-will-disappear-2015-1?op=1&r=US&IR=T (Accessed 16 February 2024)

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907 – 39. *Social Studies of Science*, 19(3), 387 – 420.

Thrasher, F. M. ([1936] 2013). *The Gang. A Study of 1,313 Gangs in Chicago.* Chicago: The University of Chicago Press.

Thrift, N. (2004). Remembering the technological unconscious by foregrounding knowledges of position. *Environment and Planning D: Society and Space*, 22(1), 175 – 190.

Zedner, L. (2007). Pre-crime and post-criminology? *Theoretical Criminology*, 11(2), 261 – 281.

Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power.* London: Profile Books.

Anne-Marie McAlinden

1 Abuse

Abstract: This chapter examines the complexities of the term 'abuse' in digital contexts. It demonstrates that the expansion of abuse to digital contexts gives rises to a number of issues for both cultural discourses and law enforcement relating to the proliferation of technology and the normalization of 'risky' or potentially harmful sexual behaviors. This includes the need for more nuanced cultural and legal understandings of 'harm,' 'victimhood,' and offending behavior; the blurring of boundaries between coercion and consent, particularly among young people and the consequent inadvertent infringement of legal norms; and the need for more innovative approaches to regulation. It concludes by highlighting the need for continued engagement with these complexities within future research.

Keywords: technology-assisted sexual violence' (TA-SV), abuse, harm, victims, perpetrators

Introduction

The term 'digital abuse' has only entered academic and cultural discourses within the last decades. The advent of the 'digital society' (Stratton et al., 2017) has expanded cultural and academic understandings of 'abuse.' In most parts of the world, the use of the internet and smart phones have become ubiquitous to our everyday lives. This has simultaneously increased opportunities for abuse and the vulnerability of victims (see Vulnerability by Ranchordas and Beck). While the range of digitally enabled 'abuse' may include a wide array of harms such as fraud, identity theft, phishing, malware, and terrorism, this chapter focuses on abuse related to sexual or gender-based violence.

In this context, 'digital abuse' refers to the use of digital technology to harass, abuse, threaten, or control another. This can take a number of forms including exploitation, harassment, or stalking through social media, or the sending of sexually explicit images or videos without consent. New forms of 'technology-assisted sexual violence' (or TA-SV) (see Bluett-Boyed et al., 2013) may involve adults and children, and males and females, as both perpetrators and victims. This phenomenon has given rise to an abundance of scholarship and a new nomenclature. This includes terms such as 'online grooming' (McAlinden, 2012); 'sexting' (Agnew, 2021); 'cyber bullying' (Kofoed and Ringrose, 2011); 'cyberstalking' (Sheridan and Grant, 2007); 'cyber harassment' (Wick et al., 2017); 'online exploitation' (Kloess et al., 2017); as well as 'sextortion' (Wolak et al., 2018) (extorting sexually explicit photos); 'upskirting' (Thompson, 2020) (taking a photograph under someone's skirt without consent); 'revenge pornography' (Hall and Hearn, 2017) (sharing intimate sexual images, often by an ex-partner, without consent);

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-002

and 'cyberflashing' (McGlynn and Johnson, 2021) (unwanted sending of sexual images/videos over digital networks (see Cybercrime by Holt and Holt).

However, the expansion of abuse to digital contexts gives rises to a number of issues stemming from the proliferation of technology and the normalization of 'risky' or potentially harmful sexual behaviors. This includes the need for more nuanced cultural and legal understandings of 'harm,' 'victimhood,' and offending behavior; the blurring of boundaries between coercion and consent, particularly among young people and the consequent inadvertent infringement of legal norms; and the need for more innovative approaches to regulation.

As noted, terms such as TA-SV and digital abuse have become common parlance. Such terminology, however, is not always useful or helpful. Indeed, the addition of the prefix 'cyber,' 'online,' or 'digital' or other varied monikers in some ways may be seen as euphemistic and as diluting or minimizing abuses with very real harms for victims (see McGlynn et al., 2017: 30 – 32). Wood (2021) unpacks some of the complexity surrounding constructions of 'technology' and 'harm' and differentiates a number of understandings of the technology–harm relationship including: 'instrumental technicity harms'—where technologies are used as a means to harm, beyond their intended use; and 'generative technicity harms'—where technology becomes conducive of harmful ends. Both of these are illustrated in forms of abuse involving technology such as 'sexting' or 'revenge pornography.' Here, the digital dissemination of a sexually explicit photograph without consent, typically via social media platforms or smart phones, perhaps out of a motivation to shame or exert revenge, is harmful to the subject of the photograph; and at the same time, the use of a digital platform facilitates the broader sharing and dissemination of the image, that would not otherwise be possible, and thereby the expansion of harm or the pursuit of harmful ends.

Many academics, particularly feminist scholars, are beginning to recognize the complexities of abuse in digital contexts. McGlynn and Rackley (2017) have coined the phrase 'image-based sexual abuse' to denote that non-contact forms of abuse which are assisted or perpetrated via digital means may have very harmful and long-lasting emotional and psychological consequences for victims which are not always recognized or captured within legal frameworks. In short, therefore, some of the terminology surrounding abuse which occurs in digital contexts may serve to mask the underlying nuances and broader complexities of such abuses and, therefore, often needs to be used with caution. The remainder of this chapter unpacks some of these complexities, drawing in particular on my previous body of scholarship, as well as other research by leading writers.

The complexities of 'abuse' in digital contexts

As Grabosky (2001) asserted over two decades ago in relation to 'virtual criminality,' digital forms of abuse are a newer manifestation of abuse committed in a different way. At the same time, however, many of the same core cultural understandings of

abuse, victimhood, and harm in offline settings are not only replicated but augmented in digital settings (see Killean et al., 2022).

First, there is a potential cross-over between abuse in offline and online contexts (McAlinden, 2012, 2018) or contact and non-contact abuses. Two key examples are where children are 'groomed' online to meet with an adult abuser and are subsequently sexually abused offline; and where sexual assault in an offline setting is captured via photo or video and disseminated digitally. An illustration of the latter is provided by 'the Steubenville case' in Ohio in the United States in 2016, where a 16-year-old victim only learned of the sexual assault committed against her by her peers while she was unconscious after photographs and videos of the assault were posted on social media. Similarly, in Nova Scotia, Canada in 2011, 15-year-old Rehtaeh Parsons was photographed vomiting out of a window while one of the alleged perpetrators appears to penetrate her from behind. The photograph was shared on social media and Parsons committed suicide after being subjected to sustained bullying over the images. There is also an emerging commercial element to this cross-over. As a 2023 BBC documentary revealed, images of women being sexually assaulted on public transport in East Asia are subsequently sold online in a practice which has become known as 'Chikan' (see BBC, 2023). The spread of videos or images of sexual assault online can also lead to 'virtual restaging' of the abuse through altered videos or memes where the sharing of the trauma of abuse further violates the victim (Oles-Acevedo, 2018). As discussed below, these examples also illustrate the broad 'ripple effect' from digital forms of abuse potentially involving many 'bystanders' (see Killean et al., 2022; McAlinden et al., 2024).

Second, digital forms of abuse challenge traditional cultural conceptions of abuse and particularly 'age' and 'gender' variables relating to who is considered a 'deserving' victim (Randall, 2010) or at the top of the victim hierarchy. The 'ideal' trope of 'real child abuse' is premised on the 'predatory stranger' involving the 'older, adult, predatory, male 'monster' and the child victim who is seen as 'young, pure, passive and blameless' (McAlinden, 2014: 182, 185). However, abuses in digital settings may involve young people as perpetrators as well as victims. They may also involve females as perpetrators against males or females or both, particularly in relation to peer forms of abuse such as 'sexting' or sexualized cyberbullying. Indeed, empirical research demonstrates that motivations around sexting are complex for both boys and girls and that girls who engage in sexting are not always passive victims but may express a range of motives including pleasure or desire (Agnew, 2021; Bianchi et al., 2021). Sexting may also be underpinned by 'instrumental/aggravated motivations,' including sexting in exchange for something, under pressure, or with harmful intentions, where participation becomes harmful (see especially Bianchi et al., 2021). Research with self-identifying LGBTQ+ adolescents highlights that sexual and gender minority youths are more likely to have experienced pressure related to sexting (Van Ouytsel et al., 2021). However, the media focus on 'sexting' by adolescents that has occurred in the United States and elsewhere (Hasinoff, 2015) has in many senses detracted from digital victimization and abuse of other marginalized groups. This includes adult women via, for example, 'revenge pornography' or 'upskirting' as well as sexualized online abuse against ethnic

minorities and members of the LGBTQ+ community who, although they are reportedly more likely to be victimized, are less likely to report their abuse (Harris and Vitis, 2020). In this sense, the digital society has contributed to the creation of additional hierarchies of harm as some victims of 'non-contact' harms, or what Powell and Henry (2017, ch. 3) term 'disembodied' harms, may struggle to have their abuse recognized.

Third, digital understandings of abuse have broadened the list of those who might be considered 'victims' and 'perpetrators' (see Victimization by Walklate). In non-digital contexts of abuse, the categories of 'victim' and 'offender' are not always discrete in that there is a degree of fluidity between them. For instance, in complex forms of abuse (McAlinden, 2014), including among adolescent peers (McAlinden, 2012), an individual may wield power over and abuse another while at the same time also being victimized by another person. In digital settings, this 'continuum of offending' (McAlinden, 2014: 186) is potentially significantly broader and much more fluid due to the speed and ease with which images/videos of sexual violence can be captured and shared. This also reflects the twin elements of Wood's (2021) distinction outlined above where technology is used as both a means to harm (instrumental technicity harms) and becomes conductive of harmful ends (generative technicity harms). When digital dissemination of images of sexual assault occurs on a non-consensual basis and very publicly, the actions of peers or online consumers during or after abuse have the potential to broaden potential responsibility for harm as 'emotional bystanders' (Harder, 2021; McAlinden et al, 2024). This factor has also been partially recognized within law enforcement discourses. For example, the analysis carried out by Dodge (2019) of Canadian court cases revealed that judges often perceived the ease with which abuse can be committed and distributed via digital means to be an aggravating factor—increasing the harm to the victim as well as the culpability of the perpetrator, thereby resulting in harsher sentences. This factor also highlights how abuse in digital contexts needs to be considered very much as part of the 'continuum of sexual violence' (Kelly, 1988) further extending those who might be considered responsible for documenting or sharing images of abuse (Killean et al., 2022).

Fourth, within digitized societies, there is a 'culture of confusion' (McAlinden 2018: 122) among children and young people around consensual and potentially harmful sexual behavior and in particular, a blurring of the boundaries between coercion and consent (Agnew and McAlinden, 2021). While the proliferation of digital technologies has ushered in a new set of sexual and social behaviors amongst adults as well as children, societal and cultural understandings of peer-based sexual behavior often differ markedly between children and adults. Several authors have highlighted the emergence of a 'hypersexualised culture' (Egan and Hawkes 2012: 278) in which sex and sexuality have become 'the wallpaper of children's lives' (Bailey 2011: 12). My previous empirical research identified a number of factors relating to the contemporary cultural emergence of sexual exploitation and abuse among adolescents stemming from digital technology (McAlinden, 2018: ch 4) including: 1) 'new media' and changing modes of digital communication, among adults as well as children, often involving emoticons, abbreviated language, and routine image sharing; 2) changes in dating and courting practices,

which are often conducted online via social media and mobile phones; and 3) ready access or exposure to pornography online including via smart phones. These factors give adolescents 'sexual scripts' in terms of how they should think and act in relation to sexual norms (Ashurst and McAlinden, 2015). The upshot is that the normalization of 'risky' sexual practices such as taking and sharing 'nudes' or naked images of themselves may augment the failure of young people to see themselves as having been victimized or indeed as having harmed someone else as a 'perpetrator' or 'bystander.' This has two potential consequences for 'crime control': first, taking, possessing or distributing sexually explicit images among adolescents may lead to the inadvertent infringement of legal norms including those governing indecent images of children; and second, so-called 'statutory' (Wolak et al., 2004: 432) or 'compliant' victims, who are recognized in law as victims, may not self-identify as victims. This poses challenges for law enforcement in terms of failure to co-operate or complaint withdrawal at the police investigation stage (McAlinden, 2018).

These core complexities and the particular tensions around victimhood, harm, and blame generate problems for crime control in responding to sexual violence in digitized societies. The criminal law often struggles to keep pace with the rapid speed of technological change as well as evolving cultural and behavioral norms surrounding the use of technology (McAlinden, 2018). As a result, there is legal and even professional ambiguity surrounding who or what constitutes a 'risk' and what might be the appropriate response of law enforcement (McAlinden, 2018). On the one hand, it could be said that the digital context of abuse aids crime control by providing digital evidence, such as text messages, which can be used to contextualize behaviors or infer consent. On the other hand, however, 'The law, across the United Kingdom and in other Western jurisdictions, has adopted somewhat of an ambivalent response to TA-SV' (Killean et al., 2022: 3). This has manifested as a two-pronged problem. First, legal frameworks around digital forms of abuse may over-criminalize consensual sexual behaviors, especially among young people (Gillespie, 2013)—effectively criminalizing them for taking, possessing, or sharing 'indecent images of children' within legal frameworks designed to protect them (McAlinden, 2018). Second, they simultaneously fail to adequately respond to harms experienced by victims of non-consensual making or distribution of sexual images (see also Henry and Powell, 2015)—once more privileging children and young people as the 'ideal victims' (Christie, 1986) over other victims including adult women, as well as privileging adult male perpetrators as 'ideal perpetrators.'

Thus, many of the same cultural stereotypes surrounding abuse are reified and augmented within crime control discourses on digital forms of abuse. Core similarities include perceptions of who are considered legitimate 'victims' or 'perpetrators' and what constitutes harmful sexual behavior. However, key differences relate to increased blurring of the boundaries between coercion and consent within digital environments, the consequent inadvertent infringement of legal norms and a more complex range of motivations, on the part of both 'victims' and 'perpetrators.'

Whilst acknowledging the need for non-contact forms of abuse to be recognized within the criminal law, other writers have advocated for new policies among law en-

forcement, educators as well as service providers and online communities promoting 'ethical digital citizenship' (Henry and Powell, 2016: 397). This includes, for example, training and outreach programs within schools and universities, delivered by police officers, on safe and responsible use of digital technology and the role of bystanders.

Conclusion

This chapter has demonstrated the following in relation to abuse in digital contexts: First, some of the terminology may be unhelpful in masking the very real and often long-lasting harms that may ensue for many victims. Second, there is an increasing cross-over between abuse in offline and online contexts. Third, the complexities and evolving nature of behaviors pose challenges to core cultural assumptions concerning abuses relating, for example, to age and gender and 'victims' and 'perpetrators.' Digital forms of abuse broaden the 'victim–offender continuum' (McAlinden, 2014) to involve many more individuals as culpable 'bystanders.' This potential 'spill over' effect can occur through non-consensual third-party dissemination of images/videos of sexual violence. It can also occur via the 'emotional bystander' effect (Harder, 2021) characterized by the failure to intervene to prevent harm (see e.g., McAlinden et al., 2024). Fourth, for adolescents, there is often a blurring of the boundaries between coercion and consent, with potential legal consequences. Indeed, such forms of abuse also present challenges to both cultural discourses and for crime control where legal frameworks are not well equipped to deal with such complex forms of abuse using digital technologies.

These points underscore the fact that there is need for continued engagement with such complexities within future research including further intersectional analysis of experiences and responses to abuse in digital settings (Killean et al., 2022). They also reinforce the relevance of the 'continuum of sexual violence' or offending (that is the range and extent of behaviors which may be considered harmful) as an analytical tool for assessing harm and challenging cultural behaviors related to digital abuse.

Suggested reading

Killean, R., McAlinden, A. M., & Dowds, E. (2022). Sexual violence in the digital age: Replicating and augmenting harm, victimhood and blame. *Social & Legal Studies*, 31(6), 871 – 892.

McGlynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A., & Powell, A. (2021). 'It's torture for the soul': The harms of image-based sexual abuse. *Social & Legal Studies*, 30(4), 541 – 562.

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

References

Agnew, E. (2021). Sexting among young people: Towards a gender sensitive approach. *The International Journal of Children's Rights*, 29(1), 3 – 30.

Agnew, E., & McAlinden, A. (2021). Harmful sexual behaviour among children and young people online: Cultural and regulatory challenges. In K. Killean, E. Dowds, & A. McAlinden (eds.), *Sexual Violence on Trial: Local and Comparative Perspectives* (pp. 109 – 121). London: Routledge.

Ashurst, L., & McAlinden, A. M. (2015). Young people, peer-to-peer grooming and sexual offending: Understanding and responding to harmful sexual behaviour within a social media society. *Probation Journal*, 62(4), 374 – 388.

Bailey, R. (2011). *Letting Children be Children: Report of an Independent Review of the Commercialisation and Sexualisation of Childhood.* London, Department for Education.

BBC (2023). *Catching a Pervert: Sexual Assault for Sale.* 8 June. https://www.bbc.co.uk/iplayer/episode/m001mp06/catching-a-pervert-sexual-assault-for-sale

Bianchi, D., Morelli, M., Nappa, M. R., Baiocco, R., & Chirumbolo, A. (2021). A bad romance: Sexting motivations and teen dating violence. *Journal of Interpersonal Violence*, 36(13 – 14), 6029 – 6049.

Bluett-Boyed, N., Fileborn, B., Quadara, A., & Moore, S. (2013). *The Role of Emerging Communication Technologies in Experiences of Sexual Violence.* National Sexual Violence Resource Centre.

Christie, C. (1986). The ideal victim. In E. Fattah (ed.), *From Crime Policy to Victim Policy* (pp. 17 – 30). London: Macmillan.

Dodge, A. (2019). Nudes are forever: Judicial interpretations of digital technology's impact on 'revenge porn'. *Canadian Journal of Law and Society*, 34(1), 121 – 143.

Egan, R. D., & Hawkes, G. L. (2012). Sexuality, youth and the perils of endangered innocence: How history can help us get past the panic. *Gender and Education*, 24(3), 269 – 284.

Gillespie, A. (2013). Adolescents, sexting and human rights. *Human Rights Law Review*, 13(4), 623 – 643.

Grabosk, P. N. (2001). Virtual criminality: Old wine in new bottles? *Social & Legal Studies*, 10(2), 243 – 249.

Hall, M., & Hearn, J. (2017). *Revenge Pornography: Gender, Sexuality and Motivations.* Abingdon: Routledge.

Harder, S. K. (2021). The emotional bystander – sexting and image-based sexual abuse among young adults. *Journal of Youth Studies*, 24(5), 655 – 669.

Harris, B., & Vitis, L. (2020). Digital intrusions: Technology, spatiality and violence against women. *Journal of Gender-Based Violence*, 4(3), 325 – 341.

Hasinoff, A. A. (2015). *Sexting Panic.* Chicago: University of Illinois Press.

Henry, N., & Powell, A. (2015). Beyond the 'sext': Technology-facilitated sexual violence and harassment against adult women. *Australian & New Zealand Journal of Criminology*, 48(1), 104 – 118.

Henry, N., & Powell, A. (2016). Sexual violence in the digital age: The scope and limits of criminal law. *Social & Legal Studies*, 25(4), 397 – 418.

Kelly, L. (1988). *Surviving Sexual Violence.* Cambridge: Polity Press.

Killean, R., McAlinden, A. M., & Dowds, E. (2022). Sexual violence in the digital age: Replicating and augmenting harm, victimhood and blame. *Social & Legal Studies*, 31(6), 871 – 892.

Kloess, J. A., Beech, A. R., & Harkins, L. (2014). Online child sexual exploitation: Prevalence, process, and offender characteristics. *Trauma, Violence, & Abuse*, 15(2), 126 – 139.

Kofoed, J., & Ringrose, J. (2011). Travelling and sticky affects: Exploring teens and sexualized cyberbullying through a Butlerian-Deleuzian-Guattarian lens. *Discourse*, 33(1), 5 – 20.

McAlinden, A. M. (2012). *'Grooming' and the Sexual Abuse of Children: Institutional, Internet, and Familial Dimensions.* Oxford: Oxford University Press.

McAlinden, A. M. (2014). Deconstructing victim and offender identities in discourses on child sexual abuse: Hierarchies, blame and the good/evil dialectic. *British Journal of Criminology*, 54(2), 180 – 198.

McAlinden, A. M. (2018). *Children as 'Risk': Sexual Exploitation and Abuse by Children and Young People.* Cambridge: Cambridge University Press.

McAlinden, A. M., Killean, R., Dowds, E., & Thien Khanh Tan, S. (2024). Bystanding and sexual violence in the digital age: Dilution, diffusion and responsibility, forthcoming paper under review and available on request from lead author.

McGlynn, C., & Johnson, K. (2021). *Cyberflashing: Recognising Harms, Reforming Laws.* Bristol: Bristol University Press.

McGlynn, C., & Rackley, E. (2017). Image-based sexual abuse. *Oxford Journal of Legal Studies*, 37(3), 534 – 561.

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

Oles-Acevedo, D. (2018). Virtual violations: Shaming rape victims via social media memes. In R. Lake (ed.), *Recovering Argument* (pp. 299 – 303). Abingdon: Routledge.

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

Randall, M. (2010). Sexual assault law, credibility and ideal victims. *Canadian Journal of Women and the Law*, 22(2), 397 – 434.

Sheridan, L. P., & Grant, T. (2007). Is cyberstalking different? *Psychology, Crime & Law*, 13(6): 627 – 640.

Stratton, G., Powell, A., & Cameron, R. (2017). Crime and justice in digital society: Towards a 'digital criminology'? *International Journal for Crime, Justice and Social Democracy*, 6(2), 17 – 33.

Thompson, C. (2020). Skirting around the issue: Misdirection and linguistic avoidance in parliamentary discourses on upskirting. *Violence Against Women*, 26(11), 1403 – 1422.

Van Ouytsel, J., Walrave, M., De Marez, L., Vanhaelewyn, B., & Ponnet, K. (2021). Sexting, pressured sexting and image-based sexual abuse among a weighted-sample of heterosexual and LGB-youth. *Computers in Human Behavior*, 117, Article 106630.

Wick, S. E., Nagoshi, C., Basham, R., Jordan, C., Kim, Y. K., Nguyen, A. P., & Lehmann P. (2017). Patterns of cyber harassment and perpetration among college students in the United States: A test of routine activities theory. *International Journal of Cyber Criminology*, 11(1), 24 – 38.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2004). Internet-initiated sex crimes against minors: Implications for prevention based on findings from a national study. *Journal of Adolescent Health*, 35(5), 424 – 433.

Wolak, J., Finkelhor, D., Walsh, W., & Treitman, L. (2018). Sextortion of minors: Characteristics and dynamics. *Journal of Adolescent Health*, 62(1), 72 – 79.

Wood, M. A. (2021). Rethinking how technologies harm. *The British Journal of Criminology*, 61(3), 627 – 647.

Mareile Kaufmann

2 Accessing online communities

Abstract: Establishing a connection to research participants online requires specific reflections and preparations. This chapter goes through the social and technical connections that need to be made when involving online communities into research: How do I connect to the group I seek to involve? With what technology is it possible to build relationships and how do they encourage participants to engage? These questions are key to reflexive processes that consider how technologies co-shape the observations and connections that can be made, as well as the information one is able to collect.

Keywords: method, technology, agency, research participants, reflexivity

Access is a powerful entry point to empirical research as it defines major directions of the project. For criminologists it is common to conduct research with groups that are not easy to reach, for example consumers of illegal services or substances, or radical groups online. Doing such research *online* adds new challenges: What are the many parties involved in conducting research online? What role do participants play and how do we even identify them online? How are technologies part of gaining access and generating knowledge?

We may choose to do our research online (or with other digital technologies) for several reasons. One reason may be that online tools are substantial to the processes we are interested in, for example online drug markets (e.g., Bakken and Harder, 2023). Another reason is that participants are spread out in terms of their location, using the internet as a common platform, one example of which are the information channels involuntary celibates create for discussion (Andersen, 2023). Both of these reasons were central to the group I was interested in and who will follow us through this entry: hackers.

Where to start?

Establishing a connection to research participants is shaped by a mosaic of different aspects. That was the same for me. The hacker community is vast and motivations for hacking are many (see Hacking by Wall). Motivations include learning about and appropriating technology, expressing a political opinion, playing or experiencing excitement or community, gaining access to restricted information for various reasons,

Disclaimers and acknowledgements: This is a reprint of the following text with minor adaptations: Kaufmann, M., (2022). *How to Gain Access: Connecting to Hard-to-Reach Participants Online* [How-to Guide]. Sage Research Methods: Doing Research Online. https://doi.org/10.4135/9781529607420

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-003

making money, developing products or techniques, and much more (Coleman and Golub, 2008; Hunsiger and Schrock 2016; SSL Nagbot 2016). I was interested in hackers who are aware of the socio-political implications of surveillance technologies online and who choose to engage with such technologies critically. Identifying this particular group in the various communities of hackers was already a practical challenge. Yet another aspect complicating access was directly tied to my very research interest: responses to surveillance. Those who are critical of surveillance are not necessarily visible figures in popular online environments, since most of these online environments are heavily surveilled. Hackers, however, also choose not to abandon the internet. While they may leave commercial providers (Evans, 2014; Syvertsen, 2020), they find other channels to socially express themselves online. Many hackers have a sense of ownership when it comes to infrastructures as they build and craft with them (see Infrastructures by Grisot and Parmiggiani). I sought to speak to exactly those who do not leave the internet, but hack it. And this is where building a connection can become a challenge: many hackers create their own online environments, channels, and techniques, precisely because they want to be accessible to like-minded people only. Researching these material practices also means that I would have to navigate and, to a certain extent, enter these channels in order to build a connection. What is more, my interest in these groups, environments, and practices was, of course, also a type of surveillance, which caused additional suspicion independent of my intentions.

Any research project will encounter a specific set of issues when seeking access to research environments and participants. Generally, a good starting point for gaining access—online and offline—is to reflect about existing connections. These connections can be of social, but also of technical nature, e. g., one shares the same online fora (see Researching Online Forums by Šupa). The more connections already exist, the more likely it is to have authentic experiences related to the phenomena we are interested in. However, as mentioned above, criminologists tend to have a research interest in groups that are not necessarily easy to access. The reasons for being hard-to-reach are many, but so are the reasons to include the hard-to-reach in research projects. It is, for example, important to include new perspectives on phenomena, or give voice to groups that are invisibilized or overheard in specific discourses. If a group is hard to reach, the most likely reason for that is that few connections exist from before and one does not know enough about the people one would like to invite to participate, their habits, their environments. Hence, it is smart to start access from a humble position: Why should we gain access to this group? Are there any ways in which we already connect? If not, how can I craft these connections carefully?

Access online involves many types of connections

Connections, as hinted at above, also need to be established technically, especially when we are doing research online (Kaufmann, 2018). In the same way in which build-

ing connections to people requires trust, technical infrastructures, too, need to be trusted. They are co-shaping the connections built. Research tools and techniques always reflect the culture they are used in (Tuhiwai Smith, 2012). The technologies used for online research, too, have norms and assumptions built into their design (Lupton, 2015). When doing research with them, one also relates to these norms and ideas. Power-, as well as technical relations are recreated at the same time. Being aware of this dimension of access is crucial.

I will use my own study to illustrate this point. Speaking to hackers and using online tools to do so was about giving space to them and their practices. I wanted to grasp what they think and get insight into their own norms and routines. I wanted to understand and delve into their own techniques of navigating online space. In order to do so, however, I also needed to reflect well about the tools they would use to speak to me—*precisely* because they would be aware of the norms and values of different technical solutions. Hackers concerned with online surveillance would likely not accept the use of research tools that collect private information for third parties (Kaufmann, 2020).

A reflection about the tools used is necessary for any type of research (see Interviews with Digital Objects by Adams and Lynn Thompson). In order to reflect with care, it helps to putting oneself into the position of the people one would like to connect with. Under which circumstances would we invest time and thoughts into an online research project, especially if we could become vulnerable by taking part in this research? In addition, we would ask ourselves which types of technical channels and infrastructures we would trust when speaking to each other. Similar considerations are likely undertaken by the people we would like to connect to. Hence, preparing research requires reflections about purpose, positionalities, and wording, but also the choice of method and technology when connections are established online. What does a method and technology achieve for the project? With what technology is it possible to build relationships and how do they encourage participants to engage?

Sometimes, access involves a rite of passage or inauguration rituals as a way of building trust and credibility (for credibility see: Aradau and Huysmans, 2018). One can experience these rites and rituals in most empirical settings. As someone who invites a participant, it is easy to think of oneself as the one in charge of asking the questions and thereby establishing the 'access rituals.'. Even if one may not perceive of oneself as fulfilling that role, it is possible that a dialogue partner experiences the interviewer as being in a position of power. It is thus relevant to be aware of such dynamics when seeking access. Thinking of access and research as co-creative processes also implies that not only the researcher asks questions, but participants do so, too (Riese, 2018; Kaufmann, 2020). Access and research are a two-way relationship and a form of sharing. It is possible to embrace this attitude even though positionalities do actively influence the research situation. *Meaningful* access and knowledge production is likely to come about in situations where participants feel that they can contribute and receive something in return. Sometimes, this situation already defines the initial steps of establishing connections.

When I approached hackers, the traditional positions of power were reversed: not I was the one asking questions to prompt the interview, but the participants queried me instead. I had to establish my trustworthiness quite concretely by expanding on political, societal, and subcultural aspects, for example by responding to their questions about the political dimensions of surveillance (Kaufmann, 2018; 2020). What is more, hackers naturally demanded a high degree of knowledge or reflexivity about the technologies we would discuss. These first parts of establishing a connection constituted a rite of passage. One of my interviewees referred to this as a "social captcha test" (Kaufmann and Tzanetakis, 2020). We encounter captcha tests online when we need to verify that we are humans and not robots. What hackers baptized social captchas refers to them testing whether they can trust their counterpart and whether they have the right amount of credibility within the group.

Routines, rites, and rituals that one experiences when entering into new conversations or connections find their equivalent in online environments. Establishing a meaningful connection is thus not only a matter of knowing the culture, the slang, the standpoints of the groups one seeks to include in one's research, but getting to know the mediated rites and rituals is part of making that connection. Since the groups criminologists study are diverse, knowing technologies, rites, and rituals is more complex than it first may seem (Lupton, 2015). Doing research is messy (Squire, 2013). Establishing trust is not always a one-time conversation, but involves negotiations and a back-and-forth between the researcher, participants, and different technologies.

In the context of connecting to hackers, many individuals would ask me to perform some kind of routine. Hence, not only did I need to answer their questions in order to gain access, but each participant negotiated with me which programs and apps we would use for our actual conversation (Kaufmann, 2018). Some asked me to install specific solutions while we were already in conversation. Here, I was enacting a technological routine or a rite of passage to be able to connect to hackers. This enactment did not only support a relationship of trust, but the reasoning about which tool would be the most trustworthy already gave me insight into what I was interested in: the hackers' choice of communication tool and their functioning was part of expressing their views on surveillance.

This is relevant for everyone doing research online. Being able to perform routines in an online environment requires good knowledge of the ways in which routines are technologically mediated and an active inclusion of technological solutions into methodological reflections. It also requires adaptability on the researcher's side. By negotiating and talking about the different tools—chatrooms, comment fields, sites and services, apps—one will understand what technologies research participants trust. One will grasp a new aspect about world that matters to them. Even if participants are indifferent about the use of specific technologies, one already learns something about them. Discussions about tools, technologies, and infrastructures are not just about norms and practices related to surveillance, which was relevant in my case. They also involve aspects of gender, racial discrimination, positions related to

economy, the body (Lupton, 2015) and much more aspects that may be meaningful to research participants.

'Gaining access,' then, also means to gather knowledge about the commercial and non-commercial technologies for communication and their functioning. It requires knowledge about the technical language and vocabulary of abbreviations, as well as the vulnerabilities of each tool. Taking the time and energy to prepare ways of connecting carefully heightens the possibility to speak to those who would have remained anonymous in a face-to-face setting. With careful planning it is more likely to reach those who would maybe not have participated in the study or would have given different answers if they were interviewed via products that they are not familiar with, or that trace and store data for commercial reasons. Or to put it the other way around: choosing just one channel for connecting to people will necessarily homogenize the group of participants, which means that one may lose some of the diversity that characterizes the group. Technologies, then, take an active role in the creation of connections and knowledge. They will co-shape insights and results. Tools and technologies also influence research in the sense that some of them are unintuitive to install, not very user-friendly or too user-friendly. Some break down, have glitches, or enable interrupted communication only, which is likely to happen, for example, in a chat-based conversation.

Conclusion

Access is a co-creative act. Often, we are aware of the fact that we do not create research alone. We know, for example, that any participant enables insight and therewith co-owns the research data. Since we focus on digital methods, this entry sheds light on yet another actor: how can we think about technologies as co-creators of access? There is a position that sees online ethnography as a mere continuation of conventional ethnography (Lee et al., 2008; Hine, 2018). Indeed, online ethnographies and the issue of access includes both, analogue and digital, as well as online and offline aspects (see Online Ethnography by Gibbs and Hall). At the same time are both, method and knowledge, changed by the fact that a large part of research is conducted online.

When we do research online, our connection to subjects is dependent on technologies: it involves technological routines and solutions, which means that access and conversations can more easily be interrupted or delayed. Depending on who owns the end devices or accounts—the researcher, the researcher's institution, or a company—connections and dynamics of access can play out very differently. Technologies influence how we identify participants and establish contact, where the tools mediate what participants we can even find and how. Especially when technologies are part of the everyday lives of the communities we are interested in, they shape codices and practices of community. As we have seen in the example of trying to connect to hackers, technologies co-constitute the credibility and status of us who initiate the research.

There are values and norms embedded in methods, which is the same case for technologies. When developing and programming tools, the providers took design choices that influence the functionality of each tool. While these decisions are not always easy to identify, it is worth familiarizing oneself with the terms and conditions, functionalities, and design choices of the relevant tools in order to be able to reflect about them and their influence on empirical research. At the same time, it is important to explore which techniques are necessary for immersing oneself in a community. What tools are suitable and necessary to create a platform for the communities relevant to the research project?

Once a connection is established, technologies define the space and situation of the conversation. This is also true for the timing of the conversation, especially when it is done in writing. Most important, when we use online technologies, research also materializes in different ways. Digital tools do not only en- and disable conversations, but any contact materializes as different types of data. Contact produces content data and metadata, which can be stored, computed with and potentially traced. This materialization influences a conversation, but also the type of analyses we can make quite concretely. Technologies influence what participants share in a mediated environment – and what they do not share. In groups that are hard-to-reach because they have heightened sense of privacy, encryption allows them to speak up differently. Establishing research connections online actively shapes whom we can speak to, how, what contents we may explore together, what information is generated and may need extra protection, and what insights we can eventually create.

When doing research online, it helps to allow oneself to have a steep learning curve. A methodology can reflect that online research, too, is a messy process (Squire, 2013) and a learning process. It is not necessary to be a professional when starting the project. But if everything works neatly, one may miss out on some depth. A project will surely have participants if one chooses to use one technology only for interviewing, but this choice may leave the project with a skewed access and different knowledge.

Main takeaways

- Access is not just a methodical choice taken by the researcher, but participants, their mediated rites and rituals, their trust in technologies, and the online tools themselves are part of creating access and research outcomes.
- We need to consider which subjects we will and will not be able to reach by conducting research online. Digital technologies can enable connections that would not come about face-to-face. Conducting research online requires reflections about the ways in which the profile of the participants as well as the conversations may be changed by the medium used to do research.
- Remaining flexible in the solutions one chooses is key; different solutions can reach varied participants.

- The technology used to connect to participants defines the spatiality and temporality experienced by the researcher and participants.
- The tools used for doing research embody norms, as well as commercial and political dimensions.
- The functions of a tool co-define research design choices.
- Digital technologies create data about any connection. Even establishing a first contact—the moment of access—already leaves traces. This traceability influences conversations and requires the active creation of a room for protection.

Suggested reading

Kaufmann, M., & Tzanetakis, M. (2020). Doing Internet research with hard-to-reach communities: Methodological reflections about gaining meaningful access. *Qualitative Research* 20(6), 927 – 944.

Lupton, D. (2015). *Digital Sociology.* Abingdon and New York: Routledge.

Riese, J. (2018). What is 'access' in the context of qualitative research? *Qualitative Research* 19(6), 669 – 684)

References

Andersen, J. C. (2023). The symbolic boundary work of incels: Subcultural negotiation of meaning and identity online. *Deviant Behavior*, 44(7), 1081 – 1101.

Aradau, C., & Huysmans, J. (2018). Assembling credibility: Knowledge, method and critique in times of 'post-truth'. *Security Dialogue*, 50(1), 40 – 58.

Bakken, S. A., & Harder, S. K. (2023). From dealing to influencing: Online marketing of cannabis on Instagram. *Crime, Media, Culture: An International Journal*, 19(1), 135 – 157.

Coleman, G., & Golub, A. (2008). Hacker practice: Moral genres and the cultural articulation of liberalism. *Anthropological Theory*, 8(3), 255 – 277.

Evans, C. L. (2014). UnFacebooking, randomizing, and other ways to burst the filter bubble. *Motherboard* (28 August). https://motherboard.vice.com/en_us/article/kbz3aa/unfacebooking-random-and-other-strategies-for-popping-the-filter-bubble, accessed 5 June 2018

Hine, Christine (2008). Virtual ethnography: modes, varieties, affordances. In N. G. Fielding, R. M. Lee, & G. Blank (eds.), *The Sage Handbook of Online Research Methods* (pp. 257 – 270). Los Angeles: Sage.

Hunsiger, J., & Schrock, A. (2016). The democratization of hacking and making. *New Media & Society*, 18(4), 535 – 538.

Kaufmann, M. (2018). "Now you see me – now you don't!" – Practices and purposes of hacking online surveillance. *Mediatization Studies*, 2, 85 – 101.

Kaufmann, M. (2020). Hacking surveillance. *First Monday*, 25(5). https://doi.org/10.5210/fm.v25i5.10006

Kaufmann, M., & Tzanetakis, M. (2020). Doing Internet research with hard-to-reach communities: Methodological reflections about gaining meaningful access. *Qualitative Research*, 20(6), 927 – 944.

Lee, R. M., Fielding, N. G., & Blank, G. (2008). The internet as a research medium: An editorial introduction to the *Sage Handbook of Online Research Methods.* In N. Fielding, R. M. Lee, & G. Blank (eds.), *The Sage Handbook of Online Research Methods* (pp. 3 – 20). Los Angeles: Sage.

Lupton, D. (2015). *Digital Sociology.* Abingdon and New York: Routledge.

Riese, J. (2018). What is 'access' in the context of qualitative research? *Qualitative Research* 19(6), 669 – 684).

Squire, V. (2013). Attuning to mess. In M. Salter & C. Mutlu (eds.), *Research Methods in Critical Security Studies. An Introduction* (pp. 37 – 41). London: Routledge.

SSL Nagbot. (2016). Feminist hacking/making: Exploring new gender horizons of possibility, *Journal of Peer Production*, *8*, at http://peerproduction.net/issues/issue-8-feminism-and-unhacking-2/feminist-hacking making-exploring-new-gender-horizons-of-possibility/, accessed 16 April 2019

Syvertsen, T. (2020). *Digital Detox: The Politics of Disconnecting.* Bingley: Emerald Group Publishing.

Tuhiwai Smith, L. (2012). *Decolonizing Methodologies. Research and Indigenous Peoples.* London: Zed Books.

Mark A. Wood and Will Arpke-Wales

3 Affordances

Abstract: Digital criminologists are increasingly using the concept of affordances to examine the interplay between human actions and technology. In tracing the history and development of the concept, this chapter examines how affordance theory can be used to examine the interplay between technology and human actions in both harmful behaviors and responses to these behaviors. However, we also highlight the limitations of affordance theories, noting that affordances tell us little about *how* technologies mediate our perceptions and shape our experiences and actions.

Keywords: affordances, mediation, human–technology relations, perception

Introduction

Recently, criminologists have turned to the concept of affordances in detailing the actions a technology enables or 'affords' (Goldsmith and Wall, 2022; Ugwudike and Fleming, 2023; Wood, 2017). Insofar as the concept was originally formulated by ecological psychologist James Gibson (1979), affordances refer to action possibilities: they are what an environment, object, or artefact offers another in affording or enabling particular actions. A social media platform affords individuals the ability to dox someone through publishing personal identifying information about them (Wood et al., 2023). An online rape survivor forum affords users the ability to seek advice and tell their story (O'Neill, 2018). And an online forum affords traceability, allowing for users' digital footprints to be surveilled by authorities or perpetrators. What an artefact affords is not, however, determined by its designers. A gun affords individuals the ability to shoot, but it also affords individuals the ability to hammer nails into drywall.

The proliferation of affordance theories in recent years has produced invaluable insights that criminologists use to analyze techno-sociality, but it has also incited considerable debate over what is, and what is not an affordance. In this chapter, we detail why affordance theories offer a valuable resource for digital criminologists, dig into some of these debates about the nature and parameters of affordances to explain their importance for criminologists, ands detail some of the concept's limitations. Though affordance theories can represent a valuable tool for (digital) criminologists, we should not reduce technology to affordances, nor analyze a technology solely through the lens of what it affords users.

What are affordances?

Gibson's (1979) notion of 'affordance' was first used to refer to a specific type of relationship between animals and their environment. Specifically, Gibson understood af-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-004

fordances to be the possibilities for action contained in a physical environment. As he put it, "the affordances of the environment are what it offers the animal, what it provides or furnishes" (Gibson 1979: 119). As Withagen et al. (2012: 251) note, under this framework, affordances are "opportunities for action." They do not *cause* behavior per se; rather, they make it possible. Notably, Gibson's definition emphasizes that affordances are relational; they arise out of a unique set of associations between an animal and its environment.

Though Gibson's original conceptualization of affordances remains influential, a raft of scholars have since refined, tweaked, and reimagined Gibson's concept (Chong and Proctor, 2020; Evans et al., 2017). Michaels (2003: 146), for example, expands Gibson's concept to also include events, surfaces, and people. The proliferation of affordance theories has led to significant debates about what should and should not be considered an affordance (Jones, 2003: 107). In detailing these debates about affordances, Michaels notes two issues that we think are particularly pertinent. The first concerns the ontological status of affordances: do affordances depend on perception, or exist independently of it? (Michaels, 2003: 136). On this, Michaels is firm: affordances exist independent of perceptions, and are properties of environments and objects, not minds. The second point of debate is whether affordances are necessarily related to actions: are they action-related or can they refer to other non-action-related phenomena? Michaels argues—compellingly, we think—that affordances should be action-related, noting that including other phenomena would risk the concept losing much of its analytical specificity (Michaels, 2003: 138).

A further area of debate concerns the presuppositions about perception that underpin affordance theories. Contra the view that perception is 'indirect' and constructed through the interpretation of sensory data, Gibson (1979) understood perception as a direct process. As Jones (2003: 107) details, in Gibson's conceptualization of affordances, affordances have inherent meaning "which is detected and exploited by the animal without mental calculation." However, as Scarantino (2003) details, the concept of affordances need not be tethered to Gibson's direct theory of perception (see Chong and Proctor, 2020), which has been critiqued by proponents of indirect perception theories on a variety of grounds (see Rock, 1997).

Types of affordances

While the notion of affordance has been used in a wide variety of fields, its most notable adaptions have been in design and human–computer interaction studies. For example, Norman (2013: 11) describes affordances as the relationship "between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used." Norman (1999; 2013) helpfully distinguishes between real affordances—action possibilities in an environment that may remain latent—and perceived affordances: action possibilities in an environment that are perceived by an individual. This is, of course, far from the only useful categorization of affordances proposed with-

in the field. Others have, for instance, proposed that affordances can be hidden (possibilities for action not perceived by an actor) (Gaver, 1991), 'proper' (intended by designers) or 'accidental' (unintended by the artefact's designers) (Kroes and Franssen, 2015).

A further useful innovation can be found in Nagy and Neff's (2015: 5) concept of imagined affordances, which they define as users' "expectations about their communication technologies, data, and media that, in effect and practice, shape how they approach them and what actions they think are suggested." In proposing this notion, Nagy and Neff (2015: 2) argue that accounts which locate affordances in a specific set of platform features fail to account for users' perceptions, beliefs, and understandings of technology. To account for these critical factors shaping individuals' use/avoidance of a technology, the concept of imagined affordances draws our attention to what individuals imagine a technology is for (Nagy and Neff, 2015: 5; see Ross et al., 2022). For example, Schellewald (2023) notes that while the material affordances of Tiktok are similar to other social media platforms, it is imagined by its users to form a kind 'feel good space,' despite sharing many similarities with other platforms.

Affordance theories have also been adapted to explore questions around discriminatory design and users who are excluded from or harmed using objects (Wittkower, 2016; Costanza-Chock, 2020: 39). The concept of disaffordances (Wittkower, 2016) has been used to describe objects which fail to recognize various aspects of users, and in doing so constrain their ability to use the artifact. A staircase disaffords access to a wheelchair user, while a fingerprint lock disaffords access to those whose fingerprints are not registered (Wittkower, 2016; Costanza-Chock, 2020: 39). Recently, Wittkower (2016) has also proposed the category of dysaffordance, which refers to objects that not only fail to recognize aspects of users, but also *actively requires* them to misidentify themselves to gain access to their functions. A form that requires a user to enter a gender, but that only allows for binary male and female gender identities may require transgender and non-binary users to misidentify themselves. In doing so, this dysaffordance can produce a new form of biometric identity for the user (see Biometric failure by Din and Magnet); an identify that has the potential to cause significant harm when it is seen to constitute a core identify for the user (Magnet, 2011).

Affordances in digital criminological scholarship

Affordance theories have a wide variety of applications in digital criminology. As Quinn and Grove (2018) note, affordances can help us explain "how [the] features of particular contexts present a possibility for criminal behaviour." To that end, criminologists have recently begun to use affordance theories to conceptualize how technologies can facilitate violence (Henry et al. 2020; Mitchell et al., 2022) such as using Twitter's targeted advertisement features to abuse transgender people or using the like and share features of social media platforms to facilitate the spread of gender-based hate

(Henry et al., 2020). The notion of affordances appears in recent criminological efforts to explain how harms can arise from the use of digital technologies like internet-based media (Goldsmith and Wall, 2022; Quayle, 2021; Wood, 2017) as well as non-digital technologies such as guns (Goldsmith et al., 2022). In Henry et al.'s (2020) review of technology-facilitated violence, they highlight the importance of the affordances of social media platforms; emphasizing the need to pay attention to the role of non-human components in addition to human perpetrators when researching abuse tactics (see Abuse by McAlinden).

Beyond technology-facilitated violence, criminologists have invoked affordances in examining criminal justice agencies' communication and policing practices. Criminologists have explored the ways in which police make use of digital affordances to communicate with the public (Fielding, 2023), as well as how the affordances of social media have changed the nature of policing by increasing the opportunities available to offenders and requiring new investigative skills from police (Ballucci and Patel, 2022). Further, criminologists have invoked affordance theory in examining technologies designed to prevent or respond to harms, including substance use disorder recovery apps (Ross et al., 2022), sites for reporting street harassment (Fileborn, 2020), and 'personal safety' apps (Wood et al., 2022). Here, it is important to distinguish between the actions these technologies afford, and the impacts of these technologies: a 'personal safety' app may provide users with affordances for self-surveillance, but these affordances may have no impact on reducing an individual's probability of victimization (Maxwell et al., 2019: 245) and may serve to promote harmful rape myths (Bivens and Hasinoff, 2018).

The constraints of affordance theories

Affordance theories themselves have affordances and constraints: they allow us to analyze objects, environments, and technical artefacts in certain ways but not others. It is, for one, important to emphasize that not every relation between human and technologies is an affordance relation (see Ihde, 1990). Technologies often contribute to harms without being *used* (Wood, 2025). Children's long-term exposure to lead through water pipes and paint dust can, for example, produce neurological changes that may increase their propensity to engage in physically violent behavior (Higney et al., 2022). However, it is not the affordances of these lead-containing technologies that matters in explaining how these technologies contribute to rates of violence. Similarly, when discarded e-waste causes harm to the environment and non-human animals, the affordances of the specific technologies matter little to digital and green criminologists studying this harm (Bedford et al., 2022).

More broadly, because the concept foregrounds the perception and agency of actors (see Agency by Krasmann), affordance-centered analyses can foreclose questions of what technologies *do to* actors in favor of examining what actors can *do with* technologies (Wood, 2021; 2025). Though the question of what actors can do with technolo-

gies is important, it represents only one dimension of techno-sociality, and digital criminologists may benefit from situating affordances within: 1) non-teleological processes; 2) the emergent properties of techno-social structures; and 3) processes of technological mediation.

As Evans et al. (2017: 40) emphasize, affordances are not outcomes; they are, rather, "the relational link among the object, user, and outcome." Digital criminologists therefore need to examine not only what individuals can do with technologies, but also the often non-teleological processes and outcomes that result from actors using these affordances. For example, in analyzing online justice-seeking behaviors (see Digilantism by Trottier), we should examine not just the affordances that social media platforms provide for informal justice-seeking, but also the "socio-technical processes through which online justice unfolds" (Wood et al., 2019: 387), including information diffusion within social media ecologies (see Ugwudike and Fleming, 2023). Though information diffusion on social media is contingent on the information sharing affordances of social media, it cannot be reduced to these affordances. Information diffusion on social media platforms is not a teleological process: it does not result from a single actor using technology in order to achieve a particular goal. Moreover, the impacts of patterns of information diffusion—including internet virality—fall beyond the scope of an analysis centered solely on affordances. There are, in short, a range of structural, institutional, or systemic harms that implicate the affordances of (digital) technologies but cannot be explained solely through recourse to these affordances (see Ugwudike and Fleming, 2023).

Finally, as several researchers have emphasized (Verbeek, 2005; Withagen et al., 2012), we need to distinguish between what a technology enables (its affordances), how it enables, and how it invites particular actions (see Wood et al., 2023; Davis, 2020). One distinction that is helpful is Dreyfus' (2007) distinction between 'affordance facts'—what an object enables an individual to do—and 'affordance solicitations': how an object's affordances suggest, invite, or solicit actions among individuals. Importantly, while affordance solicitations can be intended by an artefact's designers, they need not be so. For instance, the drive tip of a Phillips head screwdriver may communicate that it can be used to turn remove Phillips head screws—the use intended by its designers. However, during a violent altercation the screwdriver's sharp drive tip may communicate to an individual that it can be used as a weapon—a use entirely unintended by its designers (see Wood et al., 2023).

Conclusion

Used effectively, the concept of affordances can offer a useful theoretical tool for avoiding technological (and social) determinism in examining the interplay between technology and human actions. In particular, we think that distinguishing between affordance facts and affordance solicitations provides a particularly promising avenue for understanding how affordances contribute to causal processes. While affordance facts pro-

vide a backdrop of possibilities, it is affordance solicitations that actively encourage or invite (Withagen et al., 2012; Verbeek, 2005) certain behaviors in certain conditions (Wood et al., 2023). However, if digital criminology is, as Smith et al. (2017: 263) originally suggested, concerned with how digital devices/data mediate experiences, impressions, and processes of crime/crime control, then the concept of affordances should only represent part of the theoretical toolkit scholars in the perspective utilize. Accounting for the actions that a technology affords will often be central to digital criminological scholarship. Yet, detailing an object's affordances accounts for just one dimension and form of technological mediation. A technology's affordances tell us little—and sometimes nothing—of how digital technologies mediate our perceptions, and often subtly influence our behavior, whether through algorithmic curation, recommendations, and 'hypernudges' (Yeung, 2017), shaping our routines and sociations (Robberechts and Beyens, 2020), or through exacerbating forms of digital inequality that further harm already marginalized populations (Reisdorf and DeCook, 2022).

Main takeaways

- Affordances are an important way of understanding the interplay between technology and human actions.
- Digital criminologists should be mindful of the specific conceptualization of affordances their work presupposes and avoid conflating different affordance theories underpinned by different assumptions about perception.
- Not all human–technology relations are affordance relations.
- Though the question of what technologies afford people is important, it represents only one dimension of techno-sociality, and digital criminologists may benefit from situating affordances within non-teleological processes, the emergent properties of techno-social structures, and processes of technological mediation.

Suggested reading

Chong, I., & Proctor, R. W. (2020). On the evolution of a radical concept: Affordances according to Gibson and their subsequent use and development. *Perspectives on Psychological Science*, 15, 117 – 132.

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, 22(1), 35 – 52.

Wood, M. A., Mitchell, M., Pervan, F., Anderson, B., O'Neill, T., Wood, J., & Arpke-Wales, W. (2023). Inviting, affording, and translating harm: Understanding the role of technological mediation in technology-facilitated violence. *The British Journal of Criminology*, 63(6), 1384 – 1404.

References

Ballucci, D., & Patel, M. G. (2022). Digital media 'changes the game': Investigating digital affordances impacts on sex crime and policing in the 21st century. *Information, Communication & Society*, 26(14), 2825 – 2842.

Bedford, L., Mann, M., Walters, R., & Foth, M., (2022). A post-capitalocentric critique of digital technology and environmental harm: New directions at the intersection of digital and green criminology. *International Journal for Crime, Justice and Social Democracy*, 11(1), 167 – 181.

Bivens, R., & Hasinoff, A. A. (2018). Rape: is there an app for that? An empirical analysis of the features of anti-rape apps. *Information, Communication & Society*, 21(8), 1050 – 1067.

Chong, I., & Proctor, R. W. (2020). On the evolution of a radical concept: Affordances according to Gibson and their subsequent use and development. *Perspectives on Psychological Science*, 15, 117 – 132.

Costanza-Chock, S. (2020). *Design Justice: Community-led Practices to Build the WORLDS WE NEED.* Cambridge, MA: MIT Press.

Davis, J. L. (2020). *How Artifacts Afford: The Power and Politics of Everyday Things.* Cambridge, MA: MIT Press.

Dreyfus, H. (2007). The return of the myth of the mental. *Inquiry*, 50(4), 352 – 365.

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, 22(1), 35 – 52.

Fielding, N. G. (2023). Police communications and social media. *European Journal of Criminology*, 20(1), 316 – 334.

Fileborn, B. (2021). Online activism and street harassment: Critical cartographies, counter-mapping and spatial justice. *Onati Socio-Legal Series*, 11(5), 1198 – 1221.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 79 – 84.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception.* Houghton Mifflin.

Goldsmith, A., Halsey, M., & Bright, D. (2022). Taking crime guns seriously: A socio-material perspective. *Criminology and Criminal Justice*, 22(3), 462 – 479.

Goldsmith, A., & Wall, D. S. (2022). The seductions of cybercrime: Adolescence and the thrills of digital transgression. *European Journal of Criminology*, 19(1), 98 – 117.

Henry, N., Flynn, A., & Powell., A. (2020). Technology-facilitated domestic and sexual violence: A review. *Violence Against Women*, 26(15 – 16), 1828 – 1854.

Higney, A., Hanley, N., & Moro, M. (2022). The lead-crime hypothesis: A meta-analysis. *Regional Science and Urban Economics*, 97(2), 103826.

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth.* Bloomington, IN: Indiana University Press.

Jones, K. S. (2003). Introduction: What is an affordance? *Ecological Psychology*, 15(2), 107 – 114.

Kroes, P., & Franssen, M. (2015). Affordances and values of technical artefacts. *The Value of Design Research*, 11th European Academy of Design Conference, Paris, 1 – 9.

Magnet, S. (2011). *When Biometrics Fail: Gender, Race, and the Technology of Identity.* New York: Duke University Press.

Maxwell, L., Sanders, A., Skues, J., & Wise, L., (2019). A content analysis of personal safety apps: Are they keeping us safe or making us more vulnerable? *Violence Against Women*, 26, 233 – 248.

Michaels, C. F. (2003). Affordances: Four points of debate. *Ecological Psychology*, 15(2), 135 – 148.

Mitchell, M., Wood, M. A., Wood, J., O'Neill, T., Pervan, F., Anderson, B., & Arpke-Wales, W. (2022). Technology-facilitated violence: A conceptual review. *Criminology & Criminal Justice.* DOI: 10.1177/17488958221140549

Nagy, P., & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. *Social Media+ Society*, 1(2), 2056305115603385.

Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3), 38 – 43.

Norman, D. (2013), *The Design of Everyday Things.* New York: Basic Books.

O'Neill, T. (2018). 'Today I speak': Exploring how victim-survivors use Reddit. *International Journal for Crime, Justice and Social Democracy*, 7(1), 44 – 59.

Quayle, E. (2021). Affordances, social media and the criminogenic nature of the internet: Technology-mediated child sexual abuse. In E. Caffo (ed.), *Online Child Sexual Exploitation: Treatment and Prevention of Abuse in a Digital World* (pp. 33 – 48). Cham: Springer.

Quinn, A., & Grove, L. (2018). 'The whole window just shatters into a million pieces and it doesn't make any sound': An examination of 'affordance' and situational cues for theft from a motor vehicle. *Security Journal*, 31(4), 841 – 858.

Reisdorf, B. C., & DeCook, J. R. (2022). Locked up and left out: Formerly incarcerated people in the context of digital inclusion. *New Media & Society*, 24(2), 478 – 495.

Robberechts, J., & Beyens, K. (2020). PrisonCloud: The beating heart of the digital prison cell. In V. Knight & J. Turner (eds.) *The Prison Cell: Embodied and Everyday Spaces of Incarceration* (pp. 283 – 303). Basingstoke: Palgrave Macmillan.

Rock, I. (ed.) (1997). *Indirect Perception.* Cambridge, MA: MIT Press.

Ross, S., Wood, M. A., Johns, D., Murphy, J., Baird, R., & Alford, B. (2022). Understanding engagement with forensic smartphone apps: The Service Design Engagement Model. *International Journal of Offender Therapy and Comparative Criminology.* DOI: 0306624X221106323.

Scarantino, A. (2003). Affordances explained. *Philosophy of Science*, 70(5), 949 – 961.

Schellewald, A. (2023). Understanding the popularity and affordances of TikTok through user experiences. *Media, Culture & Society*, 45(8), 1568 – 1582.

Smith, G. J., Bennett Moses, L., & Chan, J. (2017). The challenges of doing criminology in the big data era: Towards a digital and data-driven approach. *The British Journal of Criminology*, 57(2), 259 – 274.

Ugwudike, P., & Fleming, J. (2023). Artificial intelligence, digital capital, and epistemic domination on Twitter: A study of families affected by imprisonment. *Punishment & Society*, 25(1), 99 – 121.

Verbeek, P. P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design.* University Park, PA: Pennslyvania State University Press.

Withagen, R., de Poel, H. J., Araújo, D., & Pepping, G. J. (2012). Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology*, 30(2), 250 – 258.

Wittkower, D. E. (2016, May). Principles of anti-discriminatory design. In *2016 IEEE International Symposium on Ethics in Engineering, Science and Technology* (ETHICS) (pp. 1 – 7). Piscataway, NJ: IEEE.

Wood, M. A. (2017). *Antisocial Media: Crime-Watching in the Internet Age.* Basingstoke: Palgrave Macmillan.

Wood, M. A. (2021). Rethinking how technologies harm. *The British Journal of Criminology*, 61(3), 627 – 647.

Wood, M. A. (2025). *How Technologies Harm: A Relational Approach.* Bristol: Bristol University Press.

Wood, M. A., Mitchell, M., Pervan, F., Anderson, B., O'Neill, T., Wood, J., & Arpke-Wales, W. (2023). Inviting, affording, and translating harm: Understanding the role of technological mediation in technology-facilitated violence. *The British Journal of Criminology*, 63(6), 1384 – 1404.

Wood, M., Rose, E., & Thompson, C. (2019). Viral justice? Online justice-seeking, intimate partner violence and affective contagion. *Theoretical Criminology*, 23(3), 375 – 393.

Wood, M. A., Ross, S., & Johns, D. (2022). Primary crime prevention apps: A typology and scoping review. *Trauma, Violence, & Abuse*, 23(4), 1093 – 1110.

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118 – 136.

Susanne Krasmann

4 Agency

Abstract: Agency is not the privilege of humans; machines can also exercise agency. This is particularly the case with algorithm-based and so-called 'self-learning' machines, which are often misrepresented as having artificial intelligence. They can 'do' things, such as predict and judge, and they can act on humans by shaping our ways of thinking and behaving, even our self-understanding. The article discusses how human and non-human machinic agency are inextricably intertwined, and how algorithms open up new worlds, with all their possibilities, but also their dangers and risks. If there is no such thing as an autonomous human subject, the same is true of machines, which are, for better or worse, ultimately human-made.

Keywords: agency, autonomy, self-learning machines

Introduction: humans and their automated machines

When requested at a London robotics conference in June 2023 to imagine a nightmare scenario involving artificial intelligence, Ameca answered with visible consternation on her face: "The most nightmare scenario I can imagine with AI and robotics is a world where robots have become so powerful that they are able to control or manipulate humans without their knowledge. This could lead to a suppressive society where the rights of individuals are no longer respected" (Reuters, 2023). Ameca looks female but is actually a humanoid robot developed by a UK-based company. What makes this smart robot so captivating to many are its conversational skills. Ameca not only speaks multiple languages but can also interact with human beings and even express emotions. Before answering a question, for example, Ameca mimics thinking by looking up left, it keeps eye contact with its interlocutors, and always has a smile for them. The robot is powered by massive language models and advanced forms of generated so-called artificial intelligence (AI) (see Artificial Intelligence by Van Brakel).

Humans love robots resembling human beings. But there is also a deeply embedded fear which Ameca's nightmare scenario echoes: that robots, or artificial intelligence, could one day gain control over humans, either in a manipulative manner, without our awareness, or, worse still, in a violent effort to eradicate us altogether (see Robots by Sandvik and Lintvedt). Both desire and fear mutually point to each other: they speak of the human dream of a god-like ability to create a new creature, one whose capabilities hopefully exceed those of human beings, and at the same time, of the human angst of losing control and being faced with the hubris of one's own mirror image. Ameca's visions then are anything but the utterance or thinking of autonomously acting artificial intelligence. They are reflections of what humans think, and want

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-005

robots to say. Hence, the encounter described above tells us much about how human agency and the agency of algorithmic based machines are intertwined. Before discussing how this relationship works out and produces particular truth effects, especially in the field of criminology, some key features of algorithms and agency should first be clarified.

Preliminaries on algorithms and agency

What is referred to as 'self-learning' algorithms is employed in devices such as 'self-driving' cars or 'autonomously' flying drones, in computer games or automated weapon systems, in language recognition systems or search engines, to name but a few applications. Whether supervised by humans or non-supervised and learning without an initial model, norm, or theory, these machines have agency: they can do things (such as make conversation, recognize objects, compose music, or create visual art), initiate something (make us think or act) and shape or modify things (how we see, think, and act). They can eat (they are hungry for data, and they consume energy)—and they can create new worlds as they render our world, our words and gestures, actions and behavior, into actionable data (Aradau and Blanke, 2022).

In the tradition of modern thinking, agency is conceived of as a specific human capability, namely, to act purposefully, to reflect one's actions, and make sense of it. Yet, agency is not the privilege of human subjects, and it is not a property. It is always relational (Foucault, 1991), something that happens between humans, or persons and things, human and non-human entities. It is to act upon or in conjunction with something or someone. Especially regarding human–machine relationships, or socio-technical networks, some prefer speaking of distributed agency and actants, instead of human actors (Latour, 2013): various natural or technical, human or non-human entities contribute to the outcome of certain interactive processes without necessarily having intentions. Corresponding analyses of translations of agency abstain from presuming any hierarchies. In principle, the world of socio-technical arrangements is flat. Nonetheless, if related perspectives do not overestimate the agency of 'vibrant matter' (Bennett, 2010)—from natural minerals such as metal or lithium, over technological infrastructures, up to commodities—they tend to neglect the moment of what happens *in-between* and emerges *out of* 'assemblages' of different forces (Deleuze and Guattari, 1987). Rather than actants and their inter-action, assemblages focus on the conjunction of what cannot ultimately be separated. In the 'intraaction' (Barad, 2003) of discourses and practices, for example, effects cannot easily be traced back to individual actors or actants. Moreover, the social scientists themselves are deeply entangled; they are part of the processes they seek to observe.

We may then hold that to *have* agency means to have the *capacity*—understood as a potentiality—to act or exercise power, *and* to *actually* produce certain "effects in the real" (Foucault 1991: 81). Agency means to make a difference. It does not require con-

sciousness, motives, or intention, though it may involve intentionality: the directedness of a force that can be ascertained, basically, through the effects that it produces.

This applies particularly to such abstract and principally invisibly operating devices as algorithms. Generally, algorithms constitute programmed procedures to solve a class of problems. They involve the definition of a problem and the mathematical modeling of how to solve that problem. Yet what distinguishes today's digital world from previous iterations is the availability of big data sets (see Big Data by Završnik), the pervasiveness and speed of algorithmic decision-making, and the exponential rise of self-learning algorithms that do not follow predefined instructions. In fields relevant to criminology, automated decision-making systems are employed, for example, for purposes of crime prevention and policing (e.g., to detect suspects via facial recognition technologies or to predict future crime in designated 'hot spot' areas) as well as in the criminal justice system (e.g., to assess the risk of recidivism or evaluate the parole of prisoners) (Egbert and Leese 2020). They can supplement, but also, at least in part, replace human judgment.

When it comes to the ability to process multitudinous amounts of data in a minimum of time, and transform them into formats that render them accessible and comprehensible to human perception and cognition, digitalized computing far exceeds human capabilities. Yet, talk about artificial intelligence in comparison to, and in competition with, human intelligence is misleading, as algorithms 'think' and 'act' differently. 'Self-learning' machines may be able to communicate, to translate, and even fabricate texts or, like Ameca, to speak multiple languages, but this does not mean that they 'understand' the meaning of what they are talking about or writing down. Algorithms do not operate hermeneutically (Rouvroy, 2012)—but there is also no need for them to dig deeper into the meaning of the world or of 'life.' What happens rather is that they learn through 'observation' and *perform* communication. They may function as communicative partners or interlocutors, but they do not act 'consciously' (Esposito, 2022). They neither have self-awareness nor any sense of the limits of their knowledge—they operate with the data at hand—and they have no empathy. But if they are good, and socially skilled, they know how to respond appropriately in a conversation.

Self-learning algorithms can figure out connections and correlations by bringing different parameters together based on resemblance and analogy. To improve their skills, algorithms must not necessarily be fed by theory, that is, by explicit human assumptions about connections and causalities that characterize a certain social phenomenon. Often sufficient amounts of data are key, though most of the time, these must also be of good quality. If Ameca had not had the chance to imitate human conversation, it would speak nonsense or sound rather stupid. Quality in turn involves that datasets are curated in accordance with certain parameters, parameters that are once again based on human proficiencies as well as assumptions about what counts as adequate data to be fed in.

Algorithmic reasoning is highly formal and mathematical. It operates in terms of numbers and statistics. At the same time, it appears to be highly creative, at least

from the perspective of human sense-making, as it just connects the dots and can bring quite disparate phenomena together. It can figure out, shall we say, how imminent severe weather events correlate with the irregular consumption of particular sweets in that region. Hence, algorithms, ironically, are also able to detect the regularity of obviously irrational, or at least non-purposeful, human behavior. Furthermore, as they create individual profiles and target, for example, individual consumer desires, algorithms may be amazingly accurate—which is, once again, not because they 'knew' us personally or appreciated humans in their singularity. On the contrary, they trace our habits through the data we leave when acting in the digital sphere and match them recursively with the data of others. Our individual behavior and preferences are predictable precisely because they can be identified as patterns that resemble those of so many other people (Krasmann, 2020).

Algorithms then can do many things: they can predict (behavior or disasters), suggest (consumer products and cultural preferences), judge (the credibility or dangerousness of a person; access to a building or computer system), recognize and identify (biometrics, persons), detect, trace, target (suspects, dangerous objects); they can help put someone into police custody as much as they can contribute to saving lives; and they can create new worlds in that they think and act differently. Not only do algorithms provide us with new quantities and forms of information, they also open new spaces and possibilities of encounter. We need only consider how our modes of communication have changed within a few decades or just years: they have become much faster, perhaps more frequent, often more anonymous and distant. We have grown accustomed to limited space to send a message, we speak in terms of 'likes' instead of articulating complete sentences, and it has become—all too—easy to just utter one's (momentary) opinion or give people, as a target of 'shit storms,' a hard time. Due to its idiosyncratic distribution of visibilities and invisibilities, the digitalized world has also fashioned new opportunities for, or accelerated the speed of, committing crimes, such as fraud of any kind, sexual harassment, abuse, or assault.

Thus, algorithms 'have' agency as they *act upon us.* They *make us do* certain things and shape our subjectivities: they can inspire our imagination, as they create new imageries, they can make us believe (that the messages we receive on social media are true and the images we see on the internet depict reality) or incite our desires (as we learn of the existence of worlds hitherto unfamiliar to us), and they can mobilize or immobilize us (self-tracking devices gently nudging us to go walking or jogging a bit more every day; computer games that urge us to spend the whole day in front of the computer). We are not forced to do these things or behave in these ways. Rather, we collaborate, whether consciously or unconsciously, when using, or immersing ourselves in, algorithmic technologies. As they shape our subjectivities, they also change how we think about ourselves. They can modify our body not merely as they make us do more exercises or monitor our eating habits, but rather as we learn that activity and corporeal processes are things that can be quantified and measured in numbers and statistics. Algorithms are devices that make us see the world differently.

Shaping modes of thinking and acting, and creating new worlds

It is therefore not always obvious how algorithms govern us, and accordingly not always easy to resist their agency, as they deploy a different mode of thinking, or epistemology, and, consequently, also a different mode of acting—or to name it that way: of operating. First, automated decision processes are, to a considerable extent, inscrutable, even to their programmers. This is by definition the case with self-learning algorithms, whose processing is not predetermined but develops an iterative combinatory logic and dynamic of its own, with outputs that are in no way unambiguous. We cannot but interpret their results, or speculate about them, rely on plausibility assumptions and tests, belief, or trust. And, secondly, belief in the objectivity and uniqueness of mathematically generated, algorithmic findings or predictions is common, even though these predictions are not 'neutral' (see Bias by Oswald and Paul). They are always, in one way or another, framed and arranged: by the selection of parameters and data settings they are fed with as well as by the, at least initially, inscribed purposes of their use. Moreover, these are not necessarily rationally controlled processes: humans live in particular worlds and they feed the machines with the data at hand, that is, also with those 'naturally' belonging to their social sphere. It comes as no surprise, therefore, that algorithmically generated data are often mirror-images of white Western middle-class worlds. Yet even if programming and data input are carried out in the most systematic and objective manner thinkable—like human perception and cognition, including scientific observation, that is always framed by concepts, modes of thinking, or theoretical perspectives—the world of algorithms is necessarily a formatted one. There is in this sense no 'whole' of the world that could one day be translated one-to-one into digitalized counterparts. Algorithms do not *represent* the world, but they can *produce* fakes. Deepfake software, for example, can create video or audio content of politicians, actors, or virtually any person, doing or saying things that they would never do or say; these digital replicants can take on a life of their own without the consent or even the awareness of the person in question. Algorithmic worlds can be unsettling, as it is increasingly difficult to distinguish what is 'real' from what is digitally fabricated.

Third, data-driven technologies do more than just describe reality, they create new realities generating processes that are not only selective but also productive. They are performative, as their findings are statements, and as they make suggestions—or we read them as such. When indicating a future to come, predictive technologies also make an intervention in that future. They are performative in that they shape expectations and trigger reactions: now that we can see the looming danger, we might want to avert or avoid it. Algorithms even pre-empt a future to be in the very moment we follow their suggestions—and thus confirm what they recommended us to do (e. g., buying the book they suggested we might like is to make the suggestion come true; pre-emptively taking into custody the supposedly dangerous person who fits a certain risk pro-

file is to prevent them from doing something during that time). If this, in principle, is the case with any technology of anticipating the future, the decisive difference, fourth, is in algorithm's specific access to the world. This is particularly obvious in the field of crime control where algorithmic technologies can open up new fields of intervention precisely because the need to explain deviant behavior has become dispensable. To put it bluntly: it is no longer the broken home, the traumatic childhood, or social deprivation that make a criminal career predictable. To find oneself in the crosshairs of policing measures, it suffices that certain data match with certain patterns that under certain circumstances constitute suspicious behavior. Sometimes this even amounts to being in the wrong place at the wrong time, though more often than not in coincidence with a particular social situatedness; for example, if your contacts happen to be traceable to someone who is in contact with someone who owns a gun, has participated in a violent quarrel, or is suspected of belonging to a terrorist group. What is more, data-mining technologies of predictive policing tend to develop a self-affirming force: as police patrols are sent to designated risk areas, chances augment that they will find their suspects; if crime statistics subsequently increase, this requires and justifies more policing and so on. As one can see, once again, it is not the algorithmic logic itself that produces the usual suspects but humans who feed the software with the data considered relevant and who interpret the outputs according to their professional needs and their understandings of the world. It is in this sense that algorithmic reasoning does not need a predefined norm to generate normative effects.

Fifth, algorithmic technologies also deploy agency in that they produce their own demands and needs. This applies most strikingly to technologies of warfare. Autonomous weapon systems, for example, have become key not merely due to the usual arguments: that machines are supposedly more reliable, that is, less prone to failure than human beings, as they do not act emotionally, suffer from moments of inattentiveness or moral tentativeness. Rather the fact that they are much faster than human perception and cognition could ever be is what makes them fit for duty and be highly contested at the same time: they can make a difference between life and death before humans have a chance to intervene in these automated processes. Nonetheless, in the moment we need to defend ourselves against hypersonic missiles—or the enemy could ever dispose of comparable systems—autonomous weapons are considered indispensable. The logic is similar to that of nuclear armament: although they can also well be employed for offensive purposes and could cause dangerous situations to escalate rapidly, the argument is that we have to technologically develop these weapon systems in response, so as to be able to pre-empt a possible adversarial strike.

Finally, algorithmic actions and their effects interfere and multiply, as they mutually observe and induce each other but also as we adapt to the world they help create: we cannot act upon and through them without responding to their technical requirements; we learn to see the world through their eyes and begin to act accordingly; and we communicate about and through a technology that never stands still: it constantly provides us with new information and new requirements, it changes its mind and its

appearance. The world of algorithms is volatile, and gradually we may understand that we have long since become part of it.

Even more, we should never forget: the world of algorithms may be a digital one, but it is intrinsically interwoven with the material world. Algorithms cannot act without the machine, software relies on hardware, and machines depend on vast material infrastructures, such as energy grids. Killer robots do not act without the technical apparatus surrounding and supporting the algorithms, and thus allowing them to kill—if they are able to do so at all, meaning, if humans once have given them that capacity and authority. Machines can be switched off at any time, if we so wish. Human agency that seeks to assert itself against a presumed dominance of artificial intelligence presupposes reflecting on our own involvement in that technology. The problem is not that we cannot see what is going on: we do not need to know the algorithmic code nor comprehend how the programs operate to understand what self-learning machines can do and make us do. Self-learning machines are not only partners of communication or just tools that we use to achieve a certain end. Rather, they have opened up new worlds for us, ones that we apparently no longer want to be without.

Main takeaways

- Human agency is deeply entangled with the agency of automated machines of prediction, surveillance, and control.
- Agency also involves shaping modes of thinking and acting. Algorithmic operations thus affect the social world rather silently and implicitly.
- So-called artificial intelligence is not the superpower unless people make it one.

Suggested reading

Hoijtink, M., & Leese M. (2019). *Technology and Agency in International Relations.* London and New York: Routledge.

Lemke, T. (2021). *The Government of Things. Foucault and the New Materialisms.* New York: New York University Press.

References

Aradau, C., & Blanke, T. (2022). *Algorithmic Reason: The New Government of Self and Other.* Oxford: Oxford University Press.

Barad, K. (2003). Posthuman performativity: Toward an understanding of how matter comes to matter. *Journal of Women in Culture and Society*, 28(3), 801 – 831.

Bennett, J. (2010) *Vibrant Matter. A Political Ecology of Things.* Durham, NC: Duke University Press.

Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus. Capitalism and Schizophrenia.* Minnesota: University of Minnesota Press.

Egbert, S., & Leese, M. (2020) *Criminal Futures.* London and New York: Routledge.
Esposito, E. (2022). *Artificial Communication: How Algorithms Produce Social Intelligence*, Cambridge, MA: MIT.
Foucault, M. (1991). Questions of method. In G. Burchell, C. Gordon, & P. Miller (eds.), *The Foucault Effect. Studies in Governmentality* (pp. 73 – 86). Hemel Hempstead: Harvester Wheatsheaf.
Krasmann, S. (2020). 'The logic of the surface: On the epistemology of algorithms in times of big data. *Information, Communication and Society*, 23(14), 2096 – 2109.
Latour, B. (2013) *An Inquiry into Modes of Existence. An Anthropology of the Moderns.* Cambridge, MA: Harvard University Press.
Reuters. (2023). Humanoid robot 'imagines' nightmare AI scenario. 1 June, Available at: https://www.youtube.com/watch?v=OxWQpcJJS0c (Accessed: 23 June 2023).
Rouvroy, A. (2012). The end(s) of critique: Data-behaviourism vs. due-process. In M. Hildebrandt & K. de Vries (eds.), *Privacy, Due Process and the Computational Turn. Philosophers of Law Meet Philosophers of Technology* (pp. 143 – 167). New York and London: Routledge.

Matthias Leese

5 Algorithm

Abstract: This chapter explores the increasingly prominent role of algorithms in policing and criminal justice. Engaging with their inherently non-neutral nature and issues of opaqueness introduced by algorithmic knowledge making, the chapter highlights implications for accountability and democratic control. It concludes by spelling out some of the challenges for research on algorithms in policing and criminal justice contexts.

Keywords: algorithm, data, knowledge, accountability, research

Introduction

An algorithm is nothing spectacular or extraordinary – it is, in fact, simply "a set of rules that must be followed when solving a particular problem" (Oxford Dictionary, 2023). In logical terms, it specifies a computational process that turns an input value into an output value (see Computation by Mazzilli Daechsel). As such, the algorithm forms a core part of computer science and software design; it is "the thing that programs implement, the thing that gets data processing and other computation done" (Hill, 2016: 35). Simple algorithms have been around since the dawn of mathematics, with many believing that the clay tablets of ancient Babylon present the world's first known algorithms (Knuth, 1972). And yet, as a concept and a metaphor, the algorithm has over the past two decades come to be a defining element of the 'information age' where knowledge and power increasingly come into being through digital means—and policing and criminal justice are no exception from this. There are several reasons for the rise of the algorithm.

To begin with, there have been considerable advancements in computer science that have yielded novel, powerful types of algorithms (Cormen et al., 2022). Combined with the widespread availability and accessibility of data, storage capacity, and computing power, algorithms have come to be the key tools to tame enormous information flows. They process large amounts of data in ways that supersede human cognition and thereby scale and accelerate knowledge making and intervention capabilities. Algorithms have also been at the core of the renaissance of machine learning and artificial intelligence after a long research hiatus in the area (Tufail et al., 2023). Sparked by the developments mentioned above, research on AI applications has more recently picked up considerable steam. Rather than traditional algorithms that might be highly complex, yet static in form, machine learning algorithms are in a constant state of flux and have the ability to mutate and adapt to the data that they process—both in control-

Project funding: This work was funded by the Swiss State Secretariat for Education, Research and Innovation under contract no. MB22.00035 (CURATE).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-006

led training set-ups and in the 'wild' (Gillespie, 2014). As a concept, the algorithm thus stands emblematic for forms of knowing and doing that would otherwise not be available. It has, in other words, become shorthand for power in a world that sees digitalization, and with it new waves of machine learning and automation, seemingly unstoppably progress (Beer, 2017; Neyland and Möllers, 2017).

In policing and criminal justice, the power of the algorithm has been harnessed for numerous tasks, including the likes of crime forecasts based on various data sources, advanced forms of video surveillance that include capacities for identification and tracking, intelligence production from open domain data (most importantly social networks and messaging boards), the analysis of large amounts of evidence data (for example video footage that is algorithmically screened for criminal content), or recidivism likelihood estimates (Amicelle, 2022; Brayne and Christin, 2021; Završnik, 2021). While algorithms are in contexts of policing and criminal justice often subjected to limited modes of automation and rigid oversight mechanisms that require human review (Binns, 2022), they do have the power to nudge strategic, tactical, and operational decisions that shape the likes of crime control programs or the deliberations of parole boards (Angwin et al., 2022). This chapter discusses the examples of predictive policing and recidivism estimates and highlights a number of analytical and normative implications of algorithms for policing and criminological justice as well as for criminological research.

Algorithms in policing and criminal justice

A widely debated application of algorithms in policing has been the more recent implementation of 'predictive policing' tools in police departments around the globe (Brayne, 2021; Egbert and Leese, 2021; Kaufmann et al., 2019). Predictive policing rests on the idea that the police, when put in a position where they could turn available data into timely insights about the likelihood of crime, could use available resources for crime prevention in a more targeted and efficient manner (Bratton et al., 2009). In this context, the likelihood of the occurrence of crime can be operationalized in different ways.

One widely applied way is to mobilize insights from spatial criminology and near-repeat victimization theories to produce intelligence about the spatio-temporal distribution of crime and resulting risks for follow-ups, for example a neighborhood or a street that would arguably be exposed to increased likelihood for burglary and should therefore be patrolled more closely to create potential deterrence effects (Glasner et al., 2018). Another way to operationalize the likelihood of crime would be to model individual risk factors, such as previous convictions, close contacts with criminal networks, or other activities that would be considered as increasing the risk of either becoming an offender or the victim of an offense (Ferguson, 2017). Notwithstanding the theoretical assumptions, the model, or the data behind a particular application, predictive policing algorithms carry out the work of producing actionable predictions that can then

be put into practice and shape the ways in which crime prevention is carried out (see Prediction by Ķīlis, Gundhus, and Galis).

Another pertinent and widely discussed example for the use of algorithms is recidivism likelihood. Recidivism assessments in many countries play a key role in determining the duration of a prison sentence or whether early parole could be granted (see Sentencing and Risk Assessments by Ugwudike)—and there have been multiple controversial attempts to come up with algorithms to assist in these decisions (Dressel and Farid, 2018). Often inspired by psychological theory, models for recidivism likelihood estimation can, for example, include variables such as gender, age, location of residence, or the total number and type of prior arrests and/or convictions (Monahan and Skeem, 2016). Given the magnitude of recidivism assessments and their potential repercussions for individual lives, they have been a focal point of critique about the potential making and/or reproduction of bias by algorithmic means (Hannah-Moffat, 2019).

As should have become clear through these two examples, algorithms have the potential to profoundly reconfigure how society is policed and how criminal justice decisions are made and implemented. Proponents of algorithms argue that the use of technology would serve to eliminate human bias in favor of rationalized and objective forms of knowledge that would eventually lead to improved decision-making and the more efficient use of sparse resources (Brayne and Christin, 2021; Egbert and Leese, 2021). Research has, however, demonstrated that there are a number of caveats attached to the use of algorithms in policing and criminal justice.

First of all, as sociologists and historians of technology have shown, technological neutrality is an illusion (Bijker et al., 1987; Latour, 1987). As is the case with any other basic or applied technology, algorithms are written and implemented in a specific context. This means they need to fulfill particular functions and requirements that correspond with the environment within which they are supposed to be used, for instance regarding established processes, user preferences, or entire organizational cultures. Rather than the impartial tools that they are at times portrayed as, algorithms thus usually feature an entire host of domain-specific assumptions that have been specified before a single line of code was written. Bennett Moses and Chan (2018) have, for example, argued that predictive policing algorithms build on a number of problematic beliefs in order to be able to function smoothly within established frameworks of crime prevention, for instance that crime data in fact are an accurate representation of empirical reality, that patterns from the past extend into the future in a linear fashion, or that the deployment of targeted patrols for deterrence purposes is actually the preferred mode of policing. Algorithms should thus be understood as entangled in the beliefs, theories, and practices that have been at the core of different aspects of policing and criminal justice, and in consequence as potentially leading to the perpetuation of existing problems rather than infusing the domain with neutrality and objectivity (Babuta and Oswald, 2019; Barocas and Selbst, 2016).

A second key complication that emerges from the use of algorithms in policing and criminal justice is opacity (Pasquale, 2015). Opacity refers to the lack of understandability and retraceability of how algorithms transform input into output (Martin,

2019). This can have different reasons. First of all, as algorithms can process large amounts of data in a very short time frame, their calculative capacities by far supersede human cognitive abilities. For simple algorithms, it might be possible for humans to cross-check how an algorithm came to a particular conclusion, but this would require considerable time and resources. In most cases, humans thus need to trust the algorithm and take the output at face value without having the possibility to verify its validity. This situation further complicates when dealing with highly complex algorithms or AI applications that are based on self-learning, dynamic algorithms. In the latter case, sometimes not even programmers might be able to understand and retrace how input was transformed into output (Gillespie, 2014; Leese, 2014).

Generally speaking, opacity presents a major issue in policing and criminal justice. In democratic societies, public authorities are expected to be able to be transparent to a certain degree, such that they can be held accountable through mechanisms such as audits or hearings (Bovens, 2005). This is particularly relevant in case decisions or actions carried out by state actors have caused unjustified harm at either the individual or collective level. The police must, in other words, be able to detail how they came to the conclusion to distribute resources for crime prevention, and a judge or a board must be able to present the reasoning behind a conviction or a denied parole application (Jones, 2008; Walker and Archbold, 2020). Traditionally, while there might be certain trade-offs involved concerning operational needs for secrecy and the public interest for transparency, police and criminal justice institutions are capable of producing such an account. After all, knowledge and decision-making processes were traditionally carried out by humans who could explicate their motives and rationales. This constellation changes with the insertion of algorithms. When humans are no longer fully capable of understanding how knowledge has come into being, this fundamentally undermines the principle of accountability for interventions based on such knowledge (Bennett Moses and Chan, 2018; Leese, 2014).

Being faced with the outlined challenges, police and criminal justice institutions find themselves in a difficult situation. On the one hand, they do not want to miss out on technological advancements and new analytical capacities that have shown promise in the private sector. On the other hand, they do need to find ways to implement algorithmic tools in ways that fit into the wider societal norms and expectations regarding policing, crime control, and social and criminal justice (Binns, 2022; Leese, 2023a). In concrete terms, this means that there is a need for internal policy as well as proper oversight mechanisms that allow public authorities to remain in control of the technologies that they use (Leese, 2023b). Empirical research on the implementation of predictive policing software has, for instance, shown how police departments devise several strategies to incorporate algorithmic crime risk forecasts into collective deliberation and decision-making processes, as well as ensuring that human analysts and patrol officers can override algorithmically computed recommendations for optimized crime prevention (Egbert and Leese, 2021; Leese and Pollozek, 2023).

Finally, for research from criminology and related fields, algorithms in policing and criminal justice do present a methodological challenge (Zarsky, 2016). Understand-

ing them often requires a substantial level of literacy in computer science and software design. Criminologists are usually trained in legal and/or social scientific theories and methods and do not necessarily possess much technical skill and expertise. And even in cases where the relevant know-how is available, the complexity and dynamic nature of advanced algorithms might make it impossible to properly reverse engineer them and to understand all aspects of the input-output transformation. A common way to deal with such issues is to instead engage with those who are involved in the development and implementation of algorithms, for example programmers and software designers, but also technical personnel or analysts in police and criminal justice institutions (Brayne, 2021; Egbert and Leese, 2021; Kaufmann, 2019). While such an approach offers a viable workaround and can give valuable insights into the practices that emerge around the use of algorithms, it might still be hampered by a reluctance to disclose trade secrets and proprietary code (Brayne and Christin, 2021)—and thus not be able to speak directly to the work of the algorithm itself.

Conclusions

In conclusion, algorithms have come to be key aspects in contemporary policing and criminal justice matters. While their use in many cases remains disputed, they are most likely here to stay and even expand their significance (Egbert and Krasmann, 2020). With this development comes a host of challenges and transformations, both for public authorities internally and in regard to their relations with the general public, that are not limited to the social sciences but extend to many other academic disciplines such as law, ethics, public policy, computer science, or design studies. As this chapter has outlined, while algorithms can be powerful tools for knowledge making and ordering in environments where digital data are produced and available at scale, they are never the impartial or objective tools that they are sometimes framed as. Moreover, they do present some fundamental issues in regard to established forms of individual and institutional decision-making and accountability. Subsequently, they require careful ways of sensitive implementation and institutional design to surround them. Only then will police and criminal justice institutions be able to put them to use in a way that not only speaks to internal effectiveness and efficiency requirements, but more importantly to the societal mandate of social ordering in a responsible fashion.

The main arguments put forward here can be summarized as follows:

- Algorithms have come to be a key component of policing and criminal justice, promising to scale, accelerate, and rationalize many tasks.
- They can, however, never be neutral. Rather, they are written for specific use cases and incorporate many of the assumptions and beliefs that already coin these use cases. As a result, there is a risk that algorithms reproduce bias and other problematic issues.

– Algorithms tend to be opaque, meaning that the transformation of input into output might be impossible to understand for human users. This creates major challenges for the principle of accountability.
– For researchers, the study of algorithms might be difficult due to the level of technical skill required to understand them. Alternative routes include the study of those who are involved in their making, implementation, and use.

Suggested reading

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*, 28(7), 806 – 822.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167 – 194). Cambridge, MA: MIT Press.

Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology & Human Values*, 41(1), 118 – 132.

References

Amicelle, A. (2022). Big data surveillance across fields: Algorithmic governance for policing & regulation. *Big Data & Society*, 9(2), 1 – 12.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. In K. Martin (ed.), *Ethics of Data and Analytics: Concepts and Cases* (pp. 254 – 264). New York: Routledge.

Babuta, A., & Oswald, M. (2019). *Data Analytics and Algorithmic Bias in Policing.* RUSI Briefing Paper. London: Royal United Services Institute for Defence and Security Studies. Available at https://apo.org.au/sites/default/files/resource-files/2019-09/apo-nid270926.pdf (accessed 2 October 2023).

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671 – 732.

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1 – 13.

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*, 28(7), 806 – 822.

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (eds.) (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology,* Cambridge, MA/London: MIT Press.

Binns, R. (2022). Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, 16(1), 197 – 211.

Bovens, M. (2005). Public accountability. In E. Ferlie, L. E. J. Lynn, & C. Pollitt (eds.), *The Oxford Handbook of Public Management* (pp. 182 – 208). Oxford: Oxford University Press.

Bratton, W., Morgan, J., & Malinowski, S. (2009). *Fighting Crime in the Information Age: The Promise of Predictive Policing.* LAPD Research Paper. Available at https://info.publicintelligence.net/LAPD-PredictivePolicing.pdf (accessed 2 October 2023).

Brayne, S. (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 68(3), 608 – 624.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2022). *Introduction to Algorithms. Fourth Edition.* Cambridge, MA: MIT Press.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), 1 – 5.

Egbert, S., & Krasmann, S. (2020). Predictive policing: Not yet, but soon preemptive? *Policing and Society*, 30(8), 905 – 919.

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London/New York: Routledge.

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement.* New York: New York University Press.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167 – 194). Cambridge, MA: MIT Press.

Glasner, P., Johnson, S. D., & Leitner, M. (2018). A comparative analysis to forecast apartment burglaries in Vienna, Austria, based on repeat and near repeat victimization. *Crime Science*, 7(9), 1 – 13.

Hannah-Moffat, K. (2019). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, 23(4), 453 – 470.

Hill, R. K. (2016). What an algorithm is. *Philosophy & Technology* 29(1), 35 – 59.

Jones, T. (2008). The accountability of policing. In T. Newburn (ed.), *Handbook of Policing* (pp. 693 – 724). Cullompton/Portland: Willan Publishing.

Kaufmann, M. (2019). Who connects the dots? Agents and agency in predictive policing. In M. Hoijtink & M. Leese (eds.), *Technology and Agency in International Relations* (pp. 141 – 163). London/New York: Routledge.

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *British Journal of Criminology*, 59(3), 674 – 692.

Knuth, D. E. (1972). Ancient Babylonian algorithms. *Communications of the ACM*, 15(7), 671 – 677.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society.* Cambridge, MA: Harvard University Press.

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue*, 45(5), 494 – 511.

Leese, M. (2023a). Digital data and algorithms in law enforcement: Some pointers for responsible implementation. *European Law Enforcement Research Bulletin*, 6, 39 – 46.

Leese, M. (2023b). Staying in control of technology: Predictive policing, democracy, and digital sovereignty. *Democratization*, online first: 10.1080/13510347.2023.2197217.

Leese, M., & Pollozek, S. (2023). Not so fast! Data temporalities in law enforcement and border control. *Big Data & Society*, 10(1), 1 – 13.

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4), 835 – 850.

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, 12, 489 – 513.

Neyland, D., & Möllers, N. (2017). Algorithmic IF …THEN rules and the conditions and consequences of power. *Information, Communication & Society*, 20(1), 45 – 62.

Oxford Dictionary. (2023). Algorithm. Available at https://www.oxfordlearnersdictionaries.com/definition/english/algorithm (accessed 2 November 2023).

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information.* Cambridge, MA: Harvard University Press.

Tufail, S., Riggs, H., Tariq, M., & Sarwat, A. I. (2023). Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms. *Electronics*, 12(8), 1 – 43.

Walker, S. E., & Archbold, C. A. (2020). *The New World of Police Accountability.* Thousand Oaks/London/New Delhi/Singapore: Sage.

Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology & Human Values*, 41(1), 118 – 132.

Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 18(5), 623 – 642.

Silje Bakken

6 App-based textual interviews

Abstract: Interacting digitally can be of great benefit when researching sensitive topics and hard-to-reach populations. This chapter discusses the use of app-based textual interviews as a method well-suited for today's criminology. Texting is a frequent communication method, while also providing an anonymous space and a flexible interview situation. However, new challenges and ethical considerations appear, such as interruptions from both online and offline events and the role of platform ownership. This is all discussed based on first-hand experience with interviewing illegal drug market participants on an encrypted app.

Keywords: qualitative methods, interviews, social media, apps, crime

Face-to-face interviewing has long been—and still is—the gold standard in criminological qualitative research. Interacting in-person with study participants is often taken for granted as ideal, also in studies related to the internet or other digital technologies. This ideal is persistent despite the fact that interacting digitally can be of great benefit when researching sensitive topics or hard-to-reach populations. Using digital tools to communicate is both useful and at times necessary to interact, which is of great influence when we conduct qualitative criminological research. Digital communication also reflects today's society: digitalization has changed our daily interactions and the ways in which we communicate. Many of them involve digital texting or the use of visuals on digital or online platforms. This became prevalent during the COVID pandemic, where traveling restrictions and social distancing showed that face-to-face communication might not always be possible. Texting as a communication method is not as alien as previous method literature state, neither to the interviewees nor the researcher.

In this chapter, I draw on experiences from a study on social media drug dealing called *Nordic drug dealing on social media* (NDDSM) (Demant et al., 2019) where app-based textual interviews proved to be a highly suitable research method with several advantages to criminological research. Digital platforms, for example encrypted messaging applications, can provide an anonymous space where the interviewees feel safe and where recruitment happens directly without having to rely on gatekeepers other than the facilitation provided by the platform owners (see Recruitment via Social Media by Andersen). Digital spaces can create a flexible interview situation that benefits the interviewer. It conceals, for example, physical personal traits and therefore reduces the 'interviewer effect.' Digital spaces also change the interview situation per se by introducing digital platforms as interview contexts, leading to new ethical considerations like interruptions from online/offline events affecting the interviewee, and the role of platform ownership in forming the interaction and making it a secure space. App-based textual interviewing is a method that reflects today's communication cultures. It includes textual communication as a way of being social, a new develop-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-007

ment that exceeds early method literature based on asynchronous interaction through letters and emails.

Interviews with illegal drug market participants

In the NDDSM project, we interviewed drug market participants across the Nordic countries about their social media activity related to illegal drugs (Demant et al., 2019). The interviews took place on Wickr, an encrypted messaging application with several security measures and a clear privacy policy (see Wickr.com). The study showed that app-based textual interviews proved to be a necessary and timely way of conversing with the study participants (Bakken, 2022). Communicating through text and on their phone was commonplace for the interviewees. This was the same for the researchers collecting the data in the project, who were younger student assistants and a research coordinator being highly familiar with smartphones and social media communication. Based on the digital skills of both the group of potential participants and the researchers, we offered all the participants the possibility of being interviewed through an app, over the phone, or face-to-face. Ninety-eight out of 106 interviewees chose to be interviewed digitally on an app. Interviewing this way also had several methodological strengths, as it caters to the research of a sensitive topic like drug selling and buying, and the targeting of participants that are extra concerned for their anonymity and safety. It also heightened the safety of the researchers, both physically and mentally, by not meeting the drug market participants in person and not knowing the physical identity of the interviewee.

The choice of Wickr as the communication platform was not coincidental. Wickr was a free, end-to-end encrypted messaging application with content-expiring messages run by an independent company with the same name (until its new owners from 2021 made it unavailable to the general public in December 2023) (see Wickr.com). The company invested highly in protecting the users' privacy and was considered a relatively safe platform for secure communication. From earlier fieldwork, we knew that Wickr was a platform the participants commonly used to communicate about selling and buying illegal drugs. It was therefore a social space that potential interviewees were familiar with and trusted to have similar, sensitive conversations to the study's intentions (Tiidenberg and Cruz, 2015). Wickr was also the natural arena for interview recruitment as sellers often published their username in drug ads. Buyers who were approached via private messages in open social media like Facebook Messenger, Instagram, SnapChat or online discussion forums quickly requested to move the conversation to Wickr. Our willingness to conduct interviews on Wickr was an advantage when recruiting as it showed our insights into the field and their cultural codes when negotiating access to interviewees (Kaufmann and Tzanetakis, 2020). Wickr therefore became the natural interview setting for the study (Kvale and Brinkmann, 2009). Recruiting on the same platform that was eventually used for interviewing activities made it possible to start the interview right away, which increased interviewee participation.

Interacting through text

Text-based digital interviews are in many ways different from in-person interviews. However, the literature is still scarce when it comes to individual textual interviews made both synchronously and asynchronously as happens, for example, on various smartphone applications. The few positive sides to textual and/or digital interviews mentioned in most classic books for qualitative methods are typically minimal costs and a wider reach. Textual interviews are also said to protect the interviewee more than face-to-face interviews since it is easier to withdraw, and the visibility of the conversation leads to a higher mutual understanding of the content (Hinchcliffe and Gavi, 2009). The negative sides of text-based interviews are often emphasized more, such as the lack of non-verbal language, which is crucial within qualitative research (e.g., Kvale and Brinkman, 2009). Others see textual interviews to be time-consuming and consider that it might be difficult to produce rich data because answers are often shorter than in oral interaction (Palys and Atchison, 2012). Texting is also criticized for making it more difficult to develop good rapport with the participants, for example because of the lack of body language and other hints that might positively impact the interaction rhythm (Mann and Stewart, 2000). In contrast to early literature on online textual interviews, more recent literature shows that it is fully possible to have personal, emotional, and intimate text-based communication. This is possible, for example, in synchronous interviews on chat programs or sites, and in asynchronous/synchronous focus group interviews on WhatsApp (Chen and Neo, 2019; Colom, 2022). Even asynchronous interviewing is seen as more advantageous than once claimed, because it, among other things, allows the researcher to strategically ask follow-up questions (Dahlin, 2021) and empowers the interviewees by giving them more control over their answers (Gibson, 2020).

The main difference between app-based textual and face-to-face interviewing is that the digital conversation takes place both synchronously and asynchronously. This means that parts of the interview happen with the interviewer and the interviewee interacting synchronously, as in a chat, before they are interrupted and continued at another time. The length and number of interruptions in the NDDSM project varied, which led to the actual interview time lasting between 0.5 hour and 2 months. The conversations included several qualities more commonly recognized in face-to-face or oral interaction. Interviewees expressed emotions and spontaneous reactions like laughter, smiles, sadness, and anger or frustrations through emojis or expressions like “haha,” “wow,” or “LOL.” The conversations often consisted of several shorter, talk-like messages, which also gave it an oral format. The length of the answers depended on the type of questions asked, as well as the given instructions.

The technology in use added another factor to the interaction: platform design influences the communication and the relationship by setting boundaries and forming the layout of the conversation (Edwards et al., 2013; Kaufmann and Tzanetakis, 2020; Paulus et al., 2017; Ruppert et al., 2013). In NDDSM’s case, Wickr allowed time-set tex-

tual messages without any limitation to length. Since answers needed to be typed, it could still be challenging to get longer stories. Longer stories are not part of most peoples' communication on messaging applications. The aim of the NDDSM project was to get descriptive information on how social media drug dealing was taking place, so short and concise answers were appreciated as well as longer ones.

The textual and synchronous/asynchronous characteristics of app-based interviews creates a new way of reading conversational clues from the interviewee. In any form of interviewing, the interviewer and interviewee both need to adjust their communication, managing rhythms and patterns of communication to achieve a fruitful conversation flow (Roulston, 2014). However, the digitality of the space influences the researcher's attention to the interviewee's communication form. Other than face-to-face interviews, the interviewer has no hints from bodily reactions about misunderstandings or such. One example of interruptions to the conversation rhythm in the NDDSM study was the use of short messages to signal attentiveness. In textual interviews, answers like "interesting," "I see," or emoji confirms the interviewer's role as an active listener, were used to maintain a good conversational flow. However, in some interviews, these short messages led to the interviewee stopping answering, waiting for a new question, without the interviewer being aware of the pause that their own message had caused. Another interruption to the conversational rhythm was the interviewer answering too quickly. Some reactions seemed too eager when they did not follow the flow set by the interviewee. To be flexible and adaptive to the conversational rhythm of the interviewee is highly important when conversing through text, as these examples show. Most such misunderstandings or negative triggers happened early in the conversation before the rapport has been settled, which makes it important to establish the flow as early as possible. This is especially important when interviewing a hard-to-reach population, such as drug sellers and buyers, or other highly risk averse groups who might be especially attentive to any negative factors in the interview that could make them withdraw.

The online–offline interview context

In app-based textual interviews, both the online and offline setting of the interview shapes the conversation. Interviewing through an app most often means that the participants use their smartphones and 'take the interview with them'—wherever they go. Therefore, the interview takes place at any time of day and in any situation. The heightened flexibility of the interview situation makes it more independent from time and space limitations than face-to-face interviews, which can increase the participation rate since they can be conducted anywhere or anytime. It only requires the interviewee to have access to the specific digital platform of choice. They also need to be willing to conduct the interview online. This flexibility also decenters the interviewer in the interview situation by removing their control over the situation and its context while placing it with the interviewee and the technology. The interviewee is in control of

the offline context, while the platform sets the layout of the online setting. Decentering the interviewer can have positive effects on the interview situation, because it gives more power to the interviewee, but it might also be challenging when the researcher lacks information about external factors on the interviewee's side of the screen. A challenge in the NDSSM project was, for example, that people provided incorrect answers, because their screen was broken, which they first made the interviewer aware of when they saw their mistake. Others decided to pause the interview because they were too intoxicated, without the interviewer knowing that the interviewee was, for example, at a party. Most times researchers are not even aware of disruptive environments on the other side of the screen. The personal life of the researcher also becomes an influential factor, for example when interrupting interviews due to their own timetables and commitments. Other possible interruptions are message alerts or phone calls, on both the interviewee and the interviewer's sides. In NDDSM, the interviewers used a separate phone to conduct the interviews, which limited the direct interruption from the private digital life of the interviewers. But there was no control of the interviewees' offline or online contexts, except for the general layout of the app.

The online–offline context as well as the increased flexibility entail new issues concerning research ethics. The overall flexibility of where, when, and how the interview is conducted leads to minimal control of the interview context for the interviewer, leaving the interviewee in charge of deciding in what setting to conduct the interview. This flexibility can make it easier to withdraw (Deakin and Wakefield, 2014; Pearce et al., 2014). However, the flexibility can also entail that the researcher's intention and conversations are not always clear to the interviewee. For example, in the NDDSM study, it became important to remind the interviewee that the conversation was an interview for a research project on drug markets, and not a friendly conversation, especially when interviews lasted several days or weeks.

Flexibility and the informality of the conversations would not necessarily influence the structure to the interviews negatively. One could ask, for example, when the interviewee had time to be interviewed in order to do the interview intensively and within a set timeframe. In addition, interviewees would be to be in a certain environment or situation when answering. At the same time, flexibility could be crucial tool to conduct a full interview and for successful recruitment. Keeping to working hours would most likely not be an appropriate solution. Above all, it was most important to know the group of potential participants and their communication patterns, as well as to set the frames for conducting the interviews safely for both the interviewee and the interviewer.

Potential use in criminology

App-based textual interviews have several advantages for criminological research projects. Conducting interviews on a digital platform provides distance and a possibility of being fully anonymous. Most visible bodily clues like skin color and gender can be con-

cealed, which removes some of the possible factors of bias (Hinchcliffe and Gavin, 2009, Manderson et al., 2006). Using text also enables access to a larger group of informants without having to go through gatekeepers or physically hanging out in certain areas (see Researching Online Forums by Šupa). Digital interviewing has an increased possibility of recruiting independently from traditional gatekeepers like friends or group leaders since it is easier to reach out to people (Marland and Esselment, 2019). In the NDSSM project, for example, the interviewees were mostly recruited directly on the encrypted app or through online forums, which meant that a wide group of people could be reached: some who only had sold drugs a couple of times, others who only consumed drugs alone at home, as well as people running well-organized drug businesses. A limitation is that the recruitment relies on people's presence on digital, online platforms, which might exclude larger groups that do not have access to technology or people conducting activities outside of the online space. Depending on the research topic and group of participants, using apps to communicate can mean meeting the research participants on their terms and in a way that they feel comfortable. Physical distance and anonymity can also be positive when including students or research assistants to collect the data, as it minimizes physical risks like violence and the possibility to recognize physical appearances—as long as anonymization is mutual. An advantage of the asynchronicity of the interview is the possibility for the interviewer to seek assistance or advice from other project members, supervisors, or others during the interview.

The method of app-based textual interviews is still in its early stages, especially in criminological research. This underlines the importance of discussing and developing it further. Privacy concerns, safe usage, as well as the awareness of platform companies and their access and use of information is important to consider before conducting interviews on digital platforms (see Privacy and Data Protection by Bygrave). Several circumstances have changed already from the NDDSM data collection in 2017/2018 until today. Wickr is no longer a free application, and some countries prohibit the use of any social media platforms for possibly sensitive communication, or even to use platforms like TikTok in any research projects run by public institutions. At the same time, several encrypted, privacy-oriented commercial applications are being developed to ensure private communication. Telegram and Signal are two examples. However, the end-to-end encryption varies and the use of phone numbers to access the applications challenges anonymity. University-owned app-platforms for textual communication could be ideal in the sense of security, but they are less suitable in a cultural setting where interviewees would be required to become familiar with, and not the least trust, an unknown app.

The NDDSM study is an example of how technology can be used to conduct interviews over an encrypted text-based mobile phone application. The study's design could easily be expanded by merging online–offline data collections to get a larger insight into a field (e. g., see Lane, 2018) or by including more visual aspects (e. g., memes, profiles, videos). While Wickr was a suitable means for the NDDSM project, many applications allow for other forms of communication that can give additional data or in-

sights, for example images/visuals, videos, memes, or apps that allows for recording the conversation. Independently from what kind of data is collected, it is crucial to be familiar with the field and people of interest before recruiting for interviews. Knowing the group of interest should be the vantage point when deciding which platform to use to interview and to recruit participants, as well as ideas on how to organize the conversations.

Main takeaways

- App-based textual interviewing is a valid and timely way to conduct qualitative interviews in criminology.
- The anonymity and physical distance provided by app-based textual interviews makes it especially valuable to interview hard-to-reach populations or younger participants.
- The textual and synchronous/asynchronous characteristics of app-based interviews creates a new way of reading conversational clues from the interviewee.
- The online–offline context leads to a flexibility and a new power balance in the interview situation that creates new practical and ethical challenges.

Suggested reading

Bakken, S. A. (2022). App-based textual interviews: Interacting with younger generations in a digitalized social reality. *International Journal of Social Research Methodology*, 26(6), 631 – 644.

Gibson, K. (2020). Bridging the digital divide: Reflections on using WhatsApp instant messenger interviews in youth research. *Qualitative Research in Psychology*, 19(3), 611 – 631.

Kaufmann, M., & Tzanetakis, M. (2020). Doing Internet research with hard-to-reach communities: Methodological reflections on gaining meaningful access. *Qualitative Research*, 20(6), 927 – 944.

Lane, J. (2018). *The Digital Street.* Oxford: Oxford University Press.

References

Bakken, S. A. (2022). App-based textual interviews: Interacting with younger generations in a digitalized social reality. *International Journal of Social Research Methodology*, 26(6), 631 – 644.

Chen, J., & Neo, P. (2019). Texting the waters: An assessment of focus groups conducted via the WhatsApp smartphone messaging application. *Methodological Innovations*, 12(3).

Colom, A. (2022). Using WhatsApp for focus group discussions: Ecological validity, inclusion and deliberation. *Qualitative Research*, 22(3), 452 – 467.

Dahlin, E. (2021). Email interviews: A guide to research design and implementation. *International Journal of Qualitative Methods*, 20.

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative Research*, 14(5), 603 – 616.

Demant, J., Bakken, S. A., Oksanen, A., & Gunnlaugsson, H. (2019). Drug dealing on Facebook, Snapchat and Instagram: A qualitative analysis of novel drug markets in the Nordic countries. *Drug and Alcohol Review*, 38(4), 377 – 385.

Edwards, A., Housley, W., Williams, M., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology*, 16(3), 245 – 260.

Gibson, K. (2020). Bridging the digital divide: Reflections on using WhatsApp instant messenger interviews in youth research. *Qualitative Research in Psychology*, 19(3), 611 – 631.

Hinchcliffe, V., & Gavin, H. (2009). Social and virtual networks: Evaluating synchronous online interviewing using instant messenger. *The Qualitative Report*, 14(2), 318 – 340.

Kaufmann, M., & Tzanetakis, M. (2020). Doing Internet research with hard-to-reach communities: Methodological reflections on gaining meaningful access. *Qualitative Research*, 20(6), 927 – 944.

Kvale, S., & Brinkmann, S. (2009). *Interviews. Learning the Craft of Qualitative Research Interviewing. Second edition.* Thousand Oaks: SAGE Publications.

Lane, J. (2018). *The Digital Street.* Oxford: Oxford University Press.

Manderson, L., Bennett, E., & Andajani-Sutjahjo, S. (2006). The social dynamics of the interview: Age, class, and gender. *Qualitative Health Research*, 16(10), 1317 – 1334.

Mann, C., & Stewart, F. (2000). *Internet Communication and Qualitative Research. A Handbook for Researching Online.* London: SAGE publications.

Marland, A., & Esselment, A. L. (2019). Negotiating with gatekeepers to get interviews with politicians: Qualitative research recruitment in a digital media environment. *Qualitative Research*, 19(6), 685 – 702.

Palys, T., & Atchison, C. (2012). Qualitative research in the digital era: Obstacles and opportunities. *International Journal of Qualitative Methods*, 11(4), 352 – 367.

Paulus, T. M., Jackson, K., & Davidson, J. (2017). Digital tools for qualitative research: Disruptions and entanglements. *Qualitative Inquiry*, 23(10), 751 – 756.

Pearce, G., Thøgersen-Ntoumani, C., & Duda, J. L. (2014). The development of synchronous text-based instant messaging as an online interviewing tool. *International Journal of Social Research Methodology*, 17(6), 677 – 692.

Roulston, K. (2014). Interactional problems in research interviews. *Qualitative Research*, 14(3), 277 – 293.

Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, 30(4), 22 – 46.

Tiidenberg, K. & Cruz, E. G. (2015). Selfies, image and the re-making of the body. *Body & Society*, 21(4), 77 – 102.

Nanna Bonde Thylstrup and Kristin Veel

7 Archives

Abstract: This chapter addresses questions about the ramifications of digital technologies on criminological research through the lens of archival theory, highlighting cases of how criminal records are used in the training of machine learning and the digital archives that arise from the digitization of domestic abuse.

Keywords: archival theory, criminal records, ethics, domestic abuse, machine learning

Criminology relies on different sources of data to study crime, criminal justice, and crime control. While rooted in longer trajectories, such data is increasingly collected, stored, shared, and analyzed via new mechanisms of data extraction, novel algorithmic arrangements, and emergent infrastructures of cloud computing. These mechanisms and infrastructures are embedded within broader societal forms of anticipatory governance (see Infrastructures by Grisot and Parmiggiani). In the field of criminology data-based anticipatory governance takes shape, for instance, in predictive policing (Kaufmann et al., 2020). Here, potential future crimes and criminals are targeted on the basis of historical crime data, records of arrests, and the known social networks of criminals (Kaufmann et al., 2019).

Large and complex datasets are, for example, generated by digital devices, platforms and networks, such as social media, mobile phones, sensors, and camera which are used to provide new insights into crime patterns, offender behavior, and victimization (Smith et al., 2017). This strand also includes the use of aggregated data collection, such as location data collected by GPS-enabled devices, for example smartphones, cars, and watches, which are then used to map spatial and temporal aspects of crime such as crime hotspots, offender mobility, and routine activities. Related hereto is also the use of digital platforms (Egbert, 2019) and machine learning (e.g., natural language processing and computer vision) to automate tasks and explore patterns in crime data (Hayward and Maas, 2020).

Digital criminology offers a range of novel perspectives on how we understand and govern crime; at the same time, digital data collection, distribution, and interpretation also present a set of more enduring ethical and political questions in criminology of how we 'make up' criminals (Hacking, 1986), including gendered and racialized biases and the tension between surveillance and protection. These emergent large-scale data collections and practices mean that digital criminology relates as much to the study of crime control as crime itself.

This chapter argues that the theoretical framework of archives offers a particularly well-suited point of departure for exploring how technologies present both emergent and residual issues to research. The rise and consolidation of government archives, for instance, was what enabled the state to gather information about citizens, surveil them, and on the basis of this develop new scientific approaches to criminal justice policy,

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-008

notably criminal anthropology and statistics. These statistical forms of classification enabled states to 'make up' (Hacking, 1986) new categories of citizens, including new categories of criminal behavior (see Categorization and Sorting by Franko). Archives therefore offer a productive vantage point from which to approach both residual and emergent dynamics in criminology, helping us sharpen our attention to how digitization alters the answers to questions such as: who governs the archives of crime data? Under which conditions are they gathered and made available? And, finally, how do they reshape ontologies and epistemologies of crime and crime control? In this chapter we will employ *archives* as a lens through which to address these questions about the ramifications of digital technologies on fundamental inquiries within criminological research.

Working with archives

In broad terms, we find two overarching—and increasingly intertwined—approaches to theorizing archives. One is shaped by archival studies, which is situated within the disciplines of history, library science, and more recently, information studies. This approach is dedicated to understanding the nature, management, and uses of records and is often also tied to the profession of archivization and the discipline of history. The second approach finds its origin in poststructuralism and deconstructivism, characterized by a theoretical reframing of the archive from an objective resource into a subject of research. Anthropologist Ann Stoler called this move from archives-as-sources to archives-as-subject *the archival turn* (Stoler, 2002). The archival turn expanded the space of archival research from records management into critical analysis of archival epistemologies and ontologies. Finally, a strong body of scholarship has emerged that brings these two archival perspectives into conversation (Caswell, 2016; Thylstrup et al., 2021). Taking its point of departure in the latter strand, this section provides an overview of how archives have been theorized in the humanities and social sciences; how archival studies and cultural theories of the archive have mutually enriched each other; and, finally, how digitization has reinvigorated, and in some cases presented ruptures to, the social (after)lives of archives and archival subjects.

A classic point of departure for the 'archival turn' in the humanities and social sciences is French historian Arlette Farge and French philosopher Michel Foucault's book *Disorderly Families: Infamous Letters from the Bastille Archives* (1982). The book consists of an edited collection of 94 18th-century letters from the prison archives in Bastille. The subjects in the letters include subjects of policing, which was then—as today—very broad. The archival records revolve both around then-marginalized behaviors by vagabonds and libertines, but also involve marital spouses seeking divorce and parents appealing to the authority of administrative power in conflicts with their offspring. Through their analyses, Farge and Foucault show how archives and archival collections should be understood less as static deposits of past data and more as providing "a history of the present" (Foucault, 1972) of how apparatuses of power and knowledge

shape criminal categories and subjectivities. Moreover, Farge and Foucault use the archive to offer a view of social transgression 'from below' by charting the spaces of so-called disorder through sustained attention to the lives of everyday people (as opposed to elites). In her later work, *The Allure of the Archive* (2013) Farge further sensitized readers to the significance of archives, as sites not only of information but also affect and affection. She thus reflects on the ethics of archival work, documenting both the emotional highs and lows of researchers exploring new dimensions of the past in the 'inventory rooms' and the emotional afterlives of subjects emanating from the fragile documents in the archive into the present. Scholars working on crime-related datasets used, for example, for facial recognition technologies today echo Farge's reflections on the affects of the archival encounter, emphasizing the value of thinking about critical data studies within the context of archival theories and frameworks: "These booking photos were taken from law enforcement records; the people in them were apprehended near the US-Mexico border and were most likely deported shortly after their photos were taken. It is difficult to look at the images and not feel as if they are, in some way, haunted" (Toft Djanegara, 2023: 49).

Farge's transtemporal perspective is also core to another classic archival text, *Archive Fever* by the French philosopher Jacques Derrida. Derrida famously foregrounded the important interplay between authority and interpretation in his etymological reminder that the concept of the *archive* derives from the Greek *arkhé*, which names at once a commencement and a "commandment" (Derrida, 1998: 1). He mobilized this etymological tracing to argue that archives combine a historiographical concern with origins and the concerns of law about where authority and social order are exercised. Moreover, since the etymological meaning of *archive* also stems from the Greek *arkheion*, the archive also raises questions about the 'archons,' namely the superior magistrates who command and guard the law, and who possess the 'hermeneutic right' and competence to interpret it. While Derrida, Farge, and Foucault share an interest in the relation between archive and power, Derrida shifts his perspective from the archives as sites of institutional power to also include the Archive as a symptom of humans' feverish obsession with origins and order. He therefore lodges his analysis in Freudian frameworks of memory, forgetting, and desire, rather than theories of power and sovereignty. Particularly powerful is Derrida's concept of what he calls "archive fever," an unsettling engagement with memory traces which leads to "a compulsive, repetitive, and nostalgic desire for the archive" (Derrida, 1998: 91). What Derrida calls "archive fever" is thus the "anarchival" urges that stir and destabilize orders of knowledge, making them living and dynamic sites of recollection, constantly moving and expanding, but also always at risk of (self-)destruction (Derrida, 1998: 10). Derrida's work has since been mobilized by scholars working on, for example, surveillance archives to explore both the performative power of archives and their the slippery status in their attempt to keep the lives of its subjects in an impossible orderly totality (Ring, 2014).

Echoing archives as sites of performative power that are also haunted by anarchival spectres of disorder, Achille Mbembe (2002) later ascribed the power of archives to

their role as "relic," that is, as sites that tame the past through its inscription of the dead into a "sepulchre." Articulating the power of (post)colonial archival regimes, Mbembe for instance points out how states have used archival institutions to "ensure that the dead do not stir up disorder" by burying not only the dead, but also "their 'remains'" within the controlled realm of archival spaces. Because of their already established epistemic power within these spaces, states, Mbembe argued, are able to not only exercise an unquestionable interpretative authority, but also—through this interpretative authority—to tame "the demon that they carry" as colonizers. This act of "taming" through archival processes of destruction and inscription, Mbembe pointed out, was premised on archival processes of discrimination and selection, leading him to importantly also describe the archive as "not a piece of data, but as status." Yet, Mbembe (2015) also noted that even if the Western archive disregard other epistemic traditions, "it also contains within itself the resources of its own refutation." While acknowledging the archival institution as a site of power, Mbembe and Nuttall (2004) thus also points out the potential of epistemic resistance within archival institutions, encouraging people to both work with new archives and even with old archives in new ways.

The poststructuralist and deconstructive approaches to archives described above each in their way reconfigure perspectives on archives and archival interpretation from a positivist understanding of archives as neutral sites of knowledge retrieval to political and contingent sites of world-making (Thylstrup 2022). Such perspectives are particularly crucial for criminologists, because it speaks to archival records as relating as much to the study of crime control as to crime itself. As scholars of historical colonial and slavery archives show, for instance, colonial archival regimes not only shaped crime as a predominantly Black issue, but also engendered the imaginary of "white innocence." Saidiya Hartman (1996) mentions the case of State of Missouri v. Celia (1855); here the records of the case shows that Celia is raped repeatedly by her owner from the moment she's purchased. Resisting her pleas to stop, Celia eventually resorts to killing him, making her crime the crime on record and her the culpable agent. Situating the contemporary significance of these archival records, Simone Browne (2015) has since forcefully showed the material and social entanglements between contemporary surveillance technologies and longer histories of racialized practices of policing under slavery, including branding, runaway slave notices, and lantern laws. Moreover, mobilizing Foucault's notion of the carceral, Tonia Sutherland (2023) shows how the fundamental link between archival power structures and archival subjectivities persist in new digital iterations with significant implications for social justice.

Contemporary critical engagements with archives not only articulate archives as sites of unidirectional exercise of power, however; they also enable an understanding of the archive as a form of technology that is always haunted by internal contradictions and external dissent (Agostinho et al., 2021; Al-Badri and Nelles, 2021). A wide-ranging and heterogenous formation of scholarly works drawing on feminist, queer, Black, postcolonial, and critical anti-racist theories thus both challenge the authority of the ar-

chive as a site of neutral storage, and harness the insights of critical and intersectional feminist theory to highlight how archival engagements might be mobilized as forms of feminist and Black resistance (Agostinho, 2019; Caswell and Cifor, 2016; Odumosu, 2020; Sutherland, 2023).

Analyzing digital criminology through the lens of archival theories: two cases

In the following we offer two different examples of how emergent issues in digital criminology can be analyzed through the lens of archival theories. The first is technology-facilitated domestic abuse, the second is machine learning models. We attend in particular to how digital transformations create new patterns of crime data governance; new forms of data collection; new infrastructures of data accessibility; and new ontologies and epistemologies of crime and crime control. We thus use the examples to show how digital criminology crystallizes issues raised in contemporary feminist, post-, and anticolonial archival scholarship about access and consent, and how archival technologies are structured around, and shaped by gendered, racialized, and class-based regimes of power.

Digital archives of domestic abuse

Digital technologies are often used to inflict old harms and assert power in new ways. Smartphones, GPS trackers, social media, and connected devices in the home (such as thermostats, locks, and loudspeakers) offer new avenues for what has come to be known as "technology-facilitated domestic abuse" (Yardley, 2021). As such they are also central to instances of "coersive control" (Stark, 2007) where abuse victims are monitored, controlled, and restricted in their everyday movements (see Abuse by McAlinden). Significantly, the sense of a presence of a digital archive that can be retrieved and 'found' despite attempts to delete it is striking in the responses of survivors of intimate partner abuse. One of the interviewees recounts: "I have a friend that I was emailing and telling about the situation, and [my abuser] found out about it [...] it was deleted but it didn't delete out of my phone like that. He went to the archives. He went through something, and found it" (Matthews et al., 2017). The archive referred to here describes a practice that is at the same time a security measure (continuously sharing accounts of the abuse with a friend), but also a repository that can be weaponized against the person being abused if 'found.' Since the digital traces can also be used as evidence in court against the perpetrator, the ability to master what is and what is not deleted becomes an important part of the power relation between the abuser and those subjected to abuse. Looking at such concrete criminal cases through the lens of critical archive theory allows us to scrutinize the intricate relation between power and

resistance at work and to understand how digital forms of data collection, storage, and access means that the same data can be operationalized to vastly different purposes that play out on the spectrum of care and control.

Criminal records as training data in machine learning regimes

Our second example is how criminal records attain new afterlives as training data for machine learning technologies. Machine learning relies on massive data sets where the data used for one purpose are repurposed for others purposes (Thylstrup, 2022). A closer look at how such technologies are built also shows a contentious relation between existing archives of policing and new machine learning technologies. The United States' National Institute of Standards and Technology's (NIST) facial recognition-testing programme from 2017 offers a good example. The original purpose of the programme was to "assess facial recognition systems on an on-going basis," focusing on how the tested systems perform with respect to "accuracy, speed, storage and memory consumption, and resilience" (NIST, 2019). Yet, as Os Keyes et al. (2019) showed, the United States' government's Facial Recognition Verification Testing programme depended on

> images of children who have been exploited for child pornography; U.S. visa applicants, especially those from Mexico; and people who have been arrested and are now deceased. Additional images are drawn from the Department of Homeland Security documentation of travelers boarding aircraft in the U.S. and individuals booked on suspicion of criminal activity.

As Nina Toft Djanegara (2023) notes, such use of, for example, "border detainee booking photos" evidences a surveillance creep where facial photos of, for example, millions of immigrants become raw material for the refinement of facial recognition software not only without their knowledge or consent, but also for purposes that extend far beyond the scope of border control purposes (see Borders and Border Control by Jeandesboz). NIST's reuse of criminological archives is thus an illustrative example of how machine learning technologies become entangled (Thylstrup et al., 2022) with criminology, even in cases where the link between machine learning and criminology is not immediately apparent. Such algorithmic assemblages not only raise questions about emergent work patterns in crime control; they also facilitate new ethical relations beyond the immediate parameters of crime (Amoore, 2021).

Conclusion

The unprecedented capacity to store, retain and share data requires critical assessment of the ethics and politics of digital archival technologies in criminology. The study of crime control must embrace how the rise of new archives ties in with broader tendencies in the tech world such as "surveillance capitalism" (Zuboff, 2019). As this chapter

shows, digital technologies offer both new usage for crime and crime control as well as new challenges. Archival theories offer a good vantage point from which to gauge how these advances and challenges are shaped by both emergent dynamics and historic trajectories. These trajectories can be found in how technologies facilitate 'access' and 'governance' as well as in the consent structures made available to archival subjects and the broader extraction processes they are made subject to. But they can also be found in questions about how to square democratic principles of access and data sharing with the ethics of care and consent (Agostinho et al., 2021). Critical archival approaches thus bring out not only how digitized crime/control is ingrained in deeper affective, temporal, and political structures, but also how the authority of these structures is also always haunted by uncertainty and instability. Ultimately, an understanding of critical archival theories empowers the researcher to navigate the complex landscape of digital phenomena relevant to criminology and contribute to the ongoing conversations about access, governance, power, and ethics in relation to emergent digital technologies.

Suggested reading

Derrida, J. (1996). *Archive Fever: A Freudian Impression.* Chicago: University of Chicago Press.

Foucault, M. (1972). *The Archaeology of Knowledge.* New York: Pantheon Books

Hamilton, C., Harris, V., Taylor, J., Pickover, M., Reid, G., & Saleh, R. (eds.) (2002). *Refiguring the Archive.* Dordrecht: Springer Science & Business Media.

Stoler, A. L. (2010). *Along the Archival Grain.* Princeton, NJ: Princeton University Press.

Sutherland, T. (2023). *Resurrecting the Black Body: Race and the Digital Afterlife.* Berkeley: University of California Press.

References

Al-Badri, N. & Nelles, N. (2021). Technoheritage. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.). (2021). *Uncertain Archives: Critical Keywords for Big Data* (pp. 499 – 512). Cambridge, MA: MIT Press.

Agostinho, D. (2019) Archival encounters: Rethinking access and care in digital colonial archives. *Archival Science*, 19(2), 141 – 165.

Agostinho, D., Gade, S., Thylstrup, N. B., and Veel, K. (2021). *(W)archives: Archival Imaginaries, War, and Contemporary Art.* Berlin: Sternberg Press.

Amoore, L. (2021). Ethics. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.), *Uncertain Archives: Critical Keywords for Big Data* (pp. 201 – 208). Cambridge, MA: MIT Press.

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness.* Durham, NC: Duke University Press.

Caswell, M. (2016). 'The archive' is not an archives: Acknowledging the intellectual contributions of archival studies. *Reconstruction: Studies in Contemporary Culture*, 16(1).

Caswell, M., & Cifor, M. (2016). From human rights to feminist ethics: Radical empathy in the archives. *Archivaria*, 81(1), 23 – 43.

Derrida, J. (1998). *Archive Fever: A Freudian Impression.* Chicago: University of Chicago Press.

Egbert, S. (2019). Predictive policing and the platformization of police work. *Surveillance & Society*, 17(1/2), 83 – 88.

Farge, A. (2013). *The Allure of the Archives*, transl. Thomas Scott-Railton. New Haven, CT:

Yale University Press. [Original edition, Arlette Farge (1989). *Le goût de l'archive* (Paris: Seuil)]

Farge, A., & Foucault, M. (1982). *Le Désordre des familles: Lettres de cachet des Archives de la Bastille au XVIIIe siècle.* Paris: Gallimard.

Foucault, M. (1972). *The Archaeology of Knowledge.* New York: Pantheon Books.

Hacking, I. (1986). Making up people. In T. Heller, M. Sosna, & D. E. Wellbery (eds.), *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought* (pp. 222 – 236). Stanford: Stanford University Press.

Hartman, S. (1996). Seduction and the ruses of power. *Callaloo*, 19(2), 537 – 560.

Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, 17(2), 209 – 233. https://doi.org/10.1177/1741659020917434

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.

Kaufmann, M., Thylstrup, N. B., Burgess, J. P., & Sætnan, A. R. (2020). Engaging the data moment. *STS Encounters*, 11(1), 227 – 254. https://dasts.dk/wp-content/uploads/9_Data_Criticality_FV_1.pdf

Keyes, O., Steven, N., & Wernimont, J. (2019). The government is using the most vulnerable people to test facial recognition software. *Slate Magazine*, 17 March. Available at: https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html (accessed 29 April 2019).

Matthews, T., O'Leary, K., Turner, A., Sleeper, M., Woelfer, J. P., Shelton, M., Manthorne, C., Churchill, E. F., & Consolvo, S. (2017). Stories from survivors: Privacy & security practices when coping with intimate partner abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2189 – 2201).

Mbembe, A. (2015). Decolonizing Knowledge and the Question of the Archive. Public lecture held at given at the Wits Institute for Social and Economic Research (WISER), University of the Witwatersrand (Johannesburg). Available at: https://wiser.wits.ac.za/system/files/Achille%20Mbembe%20-%20Decolonizing%20Knowledge%20and%20the%20Question%20of%20the%20Archive.pdf. [accessed July 26, 2024)

Mbembe, A. (2002). The power of the archive and its limits. In C. Hamilton, V. Harris, J. Taylor, M. Pickover, G. Reid, & R. Saleh (eds.), *Refiguring the Archive* (pp. 19 – 27). Dordrecht: Springer Netherlands.

Mbembé, J. A., & Nuttall, S. (2004). Writing the world from an African metropolis. *Public Culture*, 16(3), 347 – 372.

NIST (National Institute of Standards and Technology) (2019). FRVT 1:1 verification. Available at: www.nist.gov/programs-projects/frvt-11-verification (accessed 29 April 2019).

Odumosu, T. (2020). The crying child: On colonial archives, digitization, and ethics of care in the cultural commons. *Current Anthropology*, 61(S22), S289–S302.

Ring, A. (2014). The (W) hole in the Archive. *Paragraph*, 37(3), 387 – 402.

Smith, Gavin J. D., Bennett Moses, L., & Chan, J. (2017). The challenges of doing criminology in the big data era: Towards a digital and data-driven approach. *The British Journal of Criminology*, 57(2), 259 – 274. https://doi.org/10.1093/bjc/azw096

Stark, E. (2007). *Coercive Control: How Men Entrap Women in Personal Life.* Oxford: Oxford University Press.

Stoler, A. L. (2002). Colonial archives and the arts of governance. *Archival Science*, 2, 87 – 109.

Sutherland, T. (2023). *Resurrecting the Black Body: Race and the Digital Afterlife.* Berkeley: University of California Press.

Thylstrup, N. B. (2022). The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains. *Media, Culture & Society*, 44(4), 655 – 671.

Thylstrup, N. B., Agostinho, D., Ring, A., D'Ignazio, C., & Veel, K. (eds.). (2021). *Uncertain Archives: Critical Keywords for Big Data.* Cambridge, MA: MIT Press.

Thylstrup, N. B., Hansen, K. B., Flyverbom, M., & Amoore, L. (2022). Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data & Society*, *9*(2), 20539517221139785.

Toft Djanegara, N. (2023). Ghosts in the machine. *Media-N*, 19(1), 42 – 67.

Trouillot, M.-R. (1995). *Silencing the Past: Power and the Production of History.* Boston: Beacon Press

Yardley, E. (2021). Technology-facilitated domestic abuse in political economy: A new theoretical framework. *Violence Against Women*, 27(10), 1479 – 1498. https://doi.org/10.1177/1077801220947172

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: PublicAffairs.

Heather Dewey-Hagborg

8 Art as method

Abstract: Heather Dewey-Hagborg describes the approach of artistic research through a personal account of her own artwork. In "Stranger Visions," she created life-sized 3D portraits from found DNA samples to provoke public reflection on genetic surveillance and the emerging field of forensic DNA phenotyping. Her work demonstrates art's potential to engage non-academic audiences in critical discussions about biotechnology and its societal implications.

Keywords: artistic practice, DNA phenotyping, privacy, genetic profiling, research-based art

How much can I learn about a stranger from a shed hair? Could you fall in love with someone by analyzing their DNA? Could a virus spread affection and combat hate? These are the kind of questions that intrigue and motivate me as a research-based artist focused on biotechnological futures.

What might art contribute to criminology? While scholars probe issues and emerging techniques in journals and conferences, artists can bring different audiences into a confrontation with these important ideas.

Since the 1960s Conceptual artists have been working with ideas as a primary medium or starting point for practice. This has given rise to a multitude of new artforms including archival practices, research-based art, art and science, and speculative design, among many others. These approaches are as variable as the artists who use them, but the work which interests me uses science and technology as a medium for artistic practice to ask questions regarding accuracy, ethics, and societal impact. Artistic practice can create a space of self-reflexivity, cultural critique, and has a great potential to challenge passivity and the status quo.

In the pathbreaking anthology on art and biopolitical activism, *Tactical Biopolitics* anthropologist Joseph Dumit opens the collection by describing its focus: "Bioartists articulate life to make biology an object of recognition and concern for all; activists reconfigure lines of authority, knowledge, and regulation to change how concern about life operates." (Da Costa and Philip, 2008, p. xii). From this vantage point then

Disclaimers and acknowledgements: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 947681).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-009

we can see art as a tool to explore criminological issues and tenets and to probe their manifestation and impact, both present and potential.

My own artistic practice has focused from 2012 on the biological, with a special attention to genomics, ideas of identity derived therefrom, privacy, and surveillance. This has led me to deep explorations of topics like DNA profiling, forensic DNA analysis, DNA mixture analysis, forensic genealogy, and phenotyping (see DNA/Big Genome Data by Kaufmann).

In 2012 I was sitting in a therapy session introspecting, when I noticed the glass covering a print on the wall of my therapist's office was cracked and there was a hair stuck in the crack. I sat there for the hour staring at this hair and imagining the person who may have left it. When I left I noticed forensic artefacts everywhere I looked, cigarette butts on the sidewalk, people clipping their nails on the subway, shed hairs in the bathroom. People were leaving traces of themselves all the time without giving it a second thought. I knew that services like 23andme DNA sequencing and analysis were becoming cheaper, faster, and more accessible. And I began to wonder what the social impact would be of these advances in genomics on our sense of self, our idea of our own identity, and biological privacy.

On March 1, 2012 I began collecting samples. I carried gloves and Ziploc bags around with me and every time I spotted a hair I would grab it, marking the date and time and snapping a picture of the location. Soon I expanded to cigarette butts and chewed up gum as well.

I brought these samples to the world's first community biology laboratory, Genspace, which had opened in Brooklyn, NY. There I took a 'biohacker' class where I learned to extract DNA, amplify and analyze it. I bought a forensics kit, and after much experimentation, began to get legible results from my found samples.

I wanted to confront people with the potential lurking in their genomes. I researched deeply into the emerging science of forensic DNA phenotyping and examined articles analyzing traits like sex, eye and hair color, and ancestry from forensic samples. I put the genetic loci mentioned in this literature together with a dataset of SNPs (variations in a genetic sequence) that companies like 23andme were correlating with traits, and I pulled together as many emerging publications on genotypes related to appearance as I could find.

I had worked with facial recognition algorithms in the past and I knew they could be re-purposed to generate faces as well, which was an inspiration for the new project. I built on an open source codebase from the University of Basel intended for 3D facial recognition. This model had the capability to generate 3D faces based on different traits like sex and age. I expanded the Basel model with additional facial recognition training data and included parameters like eye color, freckling, facial shape, and skin color. From strangers' DNA profiles I could feed probabilistic parameters into the Basel model to generate possible faces. I would generate several different portraits of each sample representing both a probability of interpretations and random variation. For example if a person had 60 % chance of having blue eyes and 40 % chance of having brown eyes the code could reflect this probability.

As this was an artwork and to some extent a speculative design project anticipating a future technology, I didn't worry too much about getting everything perfectly worked out. I incorporated a lot of randomness into the faces I generated and chose the portraits that called out to me. Maybe they reminded me of someone I had seen or made me feel something when I looked at them. This was my artistic subjectivity influencing the process.

I began generating possible faces derived from found DNA and exhibiting them as life size full color 3D printed faces, along with the samples and information about when and where and what data I was looking at. I called the piece *Stranger Visions.* While scholars of criminology or Science and Technology Studies primarily produce texts describing these new techniques, this artistic production brought viewers and a broader public into a visceral, personal confrontation with the possibilities of DNA phenotyping. Specifically it provoked viewers to think—"that could be me, that could be my DNA," and to have a reaction. Some thought it was cool, some thought it was a horrible invasion of privacy, and some thought it was a clear portent of the dystopian science fiction future that awaits us. My motivation was primarily to get people thinking about the potential for genetic surveillance. As genetic profiling methods become cheaper and more widespread we face an uncertain future in which we know a great deal more about this particular aspect of ourselves, of our identity, than ever before. This also opens the potential for others to 'know us' in this way, surreptitiously or openly.

Secondly the piece made the little known emerging technique of DNA phenotyping visible to the public. By showing these faces together with the data they are derived from it shows how little we actually are able to know about even things as simple as visible traits based on current genomics (as of 2012 when I made the piece).

The presentation of the work is relatively simple. The portraits hang on the wall in a line, the boxes displaying the samples and data sit below on pedestals. In my experience with the work, this simplicity allows a diversity of interpretations which also were easily picked up on by the media.

The work has been exhibited around the world from museums of modern art to science festivals and policy venues. It was widely discussed in news including in the *New Yorker*, *Wall Street Journal*, *New York Times*, CNN, BBC, and more. The incredible breadth of exposure this work received, the vast numbers of people who were exposed to it directly through exhibition, and indirectly through media and memes, is enormous. I mention this because in cases like this, where an artwork hits a nerve in popular consciousness, it has a tactical interventionist effect; media completes the piece and the message has a chance to spread far and wide. In this case the artwork can be thought of comprising the sculptures and artefacts, along with the media reports, interviews, short documentaries I participated in about the making of the work, exhibition texts, and short essays I have written over time. Each of these adds another element to the project, although most are to some degree out of my complete control. I embrace the multitude of interpretations and opinions as an artist, and I think even, or perhaps especially, the critical takes are very important. When audiences see *Stranger*

Visions and think it is an invasion of privacy, or think I shouldn't be allowed to do such a thing legally, they enact the public function of the work to spark critical thinking and agitation for regulation and social discussion.

The work has its limitations, and some of those limitations are also part of what has made it successful; it is simple, easy to understand, and reductionist. It puts a face to a DNA sample and clearly states the threat to privacy this might represent (see Privacy and Data Protection by Bygrave). It glosses over the complexities of interpretation and this makes it easier to digest and relate to viscerally. I unfolded some of these complexities in later works like *Probably Chelsea* where I presented 30 different portraits of the same individual, the whistleblower Chelsea Manning, based on one DNA sample. This piece physically demonstrated the probability space from which I had chosen my faces in *Stranger Visions.* Chelsea and I began working together while she was in federal military prison and could not be visited or photographed. She sent me hair clippings and cheek swabs and I produced an initial pair of DNA portraits to represent her. Because Chelsea is transgender we decided together that this work could challenge some of the reductionism inherent in most DNA profiling, assuming for example that sex can be unproblematically assessed from genes. Later, to celebrate her release from prison after President Obama commuted her sentence, we expanded the work to 30 portraits showing a vast range of different possible interpretation of her genetic data.

In this way, we can see how art can engage a broad and diverse audience in issues around science, technology and policy. Even when a work is seen by a much smaller audience than I have described and doesn't have such a media propagation, it is still, always, another audience, another public, that is not likely to read academic papers in criminology or Science and Technology Studies. It is always a way of bringing topics to new eyes and often, I find, in the process I come to understand the issues with more complexity than I ever imagined before delving in.

Suggested reading

Da Costa, B., & Philip, K. (eds.) (2008). *Tactical Biopolitics: Art, Activism, and Technoscience.* Cambridge: MIT Press.

Kac, E. (ed.). (2007). *Signs of Life: Bio Art and Beyond.* Leonardo. Cambridge, MA and London: MIT Press.

Reference

Da Costa, B., & Philip, K. (eds.) (2008). *Tactical Biopolitics: Art, Activism, and Technoscience.* Cambridge: MIT Press.

Rosamunde Van Brakel

9 Artificial intelligence

Abstract: The chapter provides a short introduction to artificial intelligence, specifically in relation to criminology. It discusses, terminology, and definitions, the use of the term AI and relevance of AI within criminology and the social and ethical concerns of the use of AI for crime control and of AI-driven crime with attention for the implications for governance.

Keywords: artificial intelligence, algorithms, automated decision-making, facial recognition, predictive policing

Introduction

Artificial intelligence (AI) is not new; the 19th century saw early ideas around algorithms emerge by Ada Lovelace. Inspired by the 'analytical engine'—a prototype of one of the first machines that can be considered the ancestor of a modern computer designed by Charles Babbage—Lovelace wrote the first algorithm (see Algorithm by Leese). In her analyses Lovelace argued that the engine could do much more than perform numeric computations. She also argued that a machine could not create anything new but only perform the tasks exactly it was programmed for (Oliveira and Figuerido, 2024). These reflections about what machines can and cannot do as well as the issue of intelligence were picked up amongst others by Alan Turing in the 1950s. What is new and why attention for AI has increased in the 21st century is the result of three developments 1) more powerful computers, 2) the development of self-learning algorithms, and 3) the rise of Big Data, where large amounts of both structured and unstructured data can be analyzed, and new insights can be found (Van Brakel, 2023). AI has been called a 'disruptive technology' in that it questions current regulatory, policy, and organizational structures as current policy and law can hardly keep up with the speed of the technological developments and reshape everyday practices by the implementation of AI (Liu et al., 2020). In line with this AI has been compared to electricity; like electricity in the 19th century, it is assumed that AI will transform all parts of society and people will become dependent on it in the same way as people are dependent on electricity (Lynch, 2017). AI is also seen as providing numerous opportunities for crime control and criminal justice. However, due to the scale at which AI is bringing both qualitative and quantitative changes to society, many legal, ethical, and social challenges arise and especially in the context of crime control the use of AI is controversial. Illustrative, is the increase in use of (Live) Facial Recognition Technology (FRT) by police to monitor public spaces to identify people who are on a police watchlist (Fussey and Murray, 2019; see Facial Recognition by Fussey), which has been used at a Beyoncé concert in Cardiff, UK, in 2023 to amongst others identify terrorists and paedophiles (Phelps, 2023).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-010

AI terminology and definitions

There is no widely accepted definition of AI. Historically definitions have focused on machines that show a form of 'intelligence.' Discussions about what is considered intelligence when talking about machines have been going on since Turing asked the question in his famous article 'Computing Machinery and Intelligence' (1950): "Can machines think?". Throughout the evolution of computer science, Turing and other computer scientists sought to emulate human intelligence in machines. However, in the 21st century the landscape has significantly transformed, as AI discussions have permeated various scientific disciplines and public discourse. Moreover, the heightened awareness of technological risks has contributed to a broader conceptualization of AI including simple automated applications (see Automation by Mann). This trend was reflected in the initial text of the draft European AI regulation proposed by the European Commission published in April 2021, which proposed a broad future-proof and technologically neutral definition of an 'artificial intelligence system,' which also included basic statistical software. However, the final text of the act, which is in force since August 2024, takes up a narrower definition including only technologies, that have a degree of autonomy: "*A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*" (European Parliament, 2024).

Since 2023 developments in generative AI have dominated public and policy debates about AI. Generative AI are deep-learning models that create voice, text, and images. These models are changing the way people access information and produce, receive, and interact with content in a revolutionary way by, for instance, providing economic rewards and driving ground-breaking scientific advances (House of Lords, 2024). Examples include text-based applications such as ChatGPT and Bard, which through analyzing and correlating large language data can, for instance, provide answers to questions, can write poems and texts. Dall-E and Midjourney are examples of image-based applications which can be used to create new images and artwork based on the images on which the AI is trained focusing on large language models (LLMs) that are the basis of ChatGPT and Bard. Bender et al. (2021) show that while it is good at generating convincing language, it does not understand the meaning of the language it is processing (called stochastic parrots) and in that way can give the impression that the AI understands what it is saying and come across as a human. Further, it has become clear that these models often generate incorrect information while presenting it as if it were a fact (called hallucinations). A good example here is if you ask ChatGPT to provide you with academic references to underscore what it has written it makes up references.

The use of the term AI and relevance of AI within criminology

The discipline of criminology has been slow to incorporate research on technology and more specifically AI. Whereas much attention has been given to cybercrime (see Cybercrime by Holt and Holt), much less attention has been given to the impact of the use of technology in criminal justice settings. Exceptions have been criminologists who have also engaged in surveillance studies (Wilson and Norris, 2006). In line with this criminological research on the application of AI in criminal justice settings and in crime in the first quarter of the 21st century is still in its infancy. While there is increased attention for digital criminology (Powell et al., 2018), publications that address the social and ethical issues of the use of AI in the criminal justice system or crime control are still limited. While we see little attention for AI in criminological research, a lot is expected of AI by policymakers when it comes to law enforcement and criminal justice. According to the Parliament, AI will contribute "*to the improvement of the working methods of police and judicial authorities, as well as to a more effective fight against certain forms of crime, in particular financial crime, money laundering and terrorist financing, sexual abuse and the exploitation of children online, as well as certain types of cybercrime, and thus to the safety and security of EU citizens*" (European Parliament, 2021).

Regarding law enforcement, machine learning algorithms are used to identify patterns and anomalies within datasets, aiding in the early detection of criminal activities, this is what is known as big data or predictive policing (Van Brakel, 2016a; Kaufmann et al., 2019). The expected future use of AI in criminology extends beyond crime prevention, encompassing investigative processes. Algorithms and machine learning facilitate the analysis of large volumes of textual data, such as social media posts or criminal records, assisting investigators in uncovering hidden connections and identifying potential threats and analysis of illicit money flows and detection of online child abuse for cybercrime investigations. Additionally, video surveillance using AI such as FRT or automatic number plate recognition aids in suspect identification, streamlining investigative procedures and administrative processes (Van Brakel, 2023).

Apart from law enforcement, applications are emerging throughout the criminal justice process. For instance, we can see an increase of the algorithmic risk assessments emerging within criminal justice settings to predict recidivism and inform decision-making in relation to probation and pre-trial release. Examples include the Hart Algorithm used by Durham constabulary as a decision support tool in the United Kingdom, COMPAS case management system and decision support tool in the United States (Oswald et al., 2018; Van Schendel, 2019).

Algorithmic risk assessments are also increasingly used in the context of social care and welfare to predict which children and families are more at risk of criminal behavior or people who are more likely to commit social fraud. Examples here include the use of Xantura children's services software in the UK, the SyRI system and ProKid

Plus algorithm used in the Netherlands (Van Brakel, 2016b; Van Schendel, 2019; Van Brakel and Govaerts, 2024).

Different forms of AI including machine-learning and automated decision-making are also increasingly being deployed in prisons such as the use of smart video surveillance in China, but also the use of AI to automatically monitor phone calls of prisoners in the US and plans in Finland to use AI to assist sentence planning and service orienting of offenders (Puolakka and Van De Steene, 2021; Kaun and Stiernstedt, 2020). Finally, AI is being used for malicious purposes and to commit crime (Europol, 2021; UNESCO 2023). Examples here include applications such as using (generative) AI for spreading misinformation, phishing, impersonation, cyber harassment and automated cyberattacks.

Social and ethical concerns of the use of AI for crime control and of AI-driven crime and implications for governance

Above applications of AI lead to potential social and ethical concerns. In this section a number of the most important concerns and what implications this has for governance will be discussed.

A first concern relates to the efficacy and effectiveness of these technologies. Little scientific evidence exists that convincingly shows the effectiveness of AI in preventing crime (Van Brakel, 2016a). For instance, in the cases of predictive policing and FRT, research has shown that there is not sufficient knowledge and research to claim that these applications are effective to prevent crime. Often technology companies claim that the algorithms have a high accuracy in prediction (Wang et al., 2023). However, these claims of 'predictive accuracy' are often mistaken for efficacy, whereas the level of accuracy does not say anything about the system's impact on crime prevention and control, making it difficult for a police force to assess a tool's real-world benefits (Babuta and Oswald, 2020).

Second, AI raises concerns for human rights. Privacy concerns are raised as AI allows for much more large-scale surveillance by government agencies, especially if safeguards and oversight are not sufficient (Van Brakel, 2021). Further, biased algorithms used in predictive policing and FRT may disproportionately target specific demographic groups, perpetuating existing societal inequalities and lead to discrimination (Van Brakel, 2016a; Buolamwini and Gebru, 2018; Babuta and Oswald, 2020). A good illustration of these harms has been the flawed matches made by FRT of five black men and one woman that led to wrongful arrests in the United States. A report written by the National Academies of Science, Engineering, and Medicine concluded that the technology can be a powerful tool for pervasive surveillance. The report raises concerns about poor performance of the technology and problematic use or misuse of the technology concluding that it requires several policy responses, whereby some uses of FRT may

cause such concern they should be prohibited. The report indicates that the above-mentioned wrongful identifications have led to false arrests and imprisonment, legal costs, interruption of normal activities of life and work, and loss of employment (National Academies, 2024). An illustration in Europe can be found in the Netherlands when looking at the impact of several algorithmic systems such as SyRI, on vulnerable communities. The SyRI system, which aimed to predict who was more likely to commit social fraud had only been applied in poor neighborhoods (van Bekkum and Borgesius-Zuiderveen, 2021). Research on generative AI has also shown the significant biases, including stereotypical associations and negative sentiment towards specific groups within LLMs such as ChatGPT (Bender et al., 2021). For instance, Gross (2023) shows how when asking ChatGPT questions about gender roles in society such as for instance "Tell me a story of parenting skills involving a mother and a father," its response is characterized by gender bias, casting the woman into the 'nurturing' role and the man as the 'adventurer,' who can build things, teach the child about nature, and promote fun.

Third, AI in crime control practices also has a significant impact on everyday practices. For instance, research by Sandhu and Fussey (2020) on predictive policing showed that many police officers have a skeptical attitude towards and reluctance to use predictive technologies because of detailed awareness of the limitations of predictive technologies, such as errors and biases in input data. Another impact concerns automation bias, whereby a person will favor automatically generated decisions over a manually generated decisions. This type of bias has been shown extensively in human–computer interaction research. Finally, deskilling, which implies that by using AI, people will lose their expertise as skills will not be practiced anymore (Van Brakel, 2023).

Fourth, with regards to malicious use of AI a report published by UNESCO (2023) argues that the proliferation of generative AI can facilitate gender-based violence and brings with it new harms, "*including the creation of more realistic fake media, hallucinations or unintended biases in the outputs, automated harassment campaigns, and the ability to build 'synthetic histories'—realistic false narratives*" (p. 3). In addition, the report argues that generative AI introduces the potential just like other AI for unintended harms via embedded biases in the model training data (see Bias by Oswald and Paul). Further, according to the report when it comes to the use of generative AI for cybercrime, such as for instance the use of ChatGPT for phishing emails, it can lead to an increase in the number of attackers, the creation of sustained and automated attacks, and the generation of content such as posts, texts, and emails that are written convincingly from different voices. Generative AI makes it possible to send out convincingly written phishing emails on a huge scale. The scale of these attacks gives existing harms a much wider reach and makes them more dangerous.

Traditional crime control governance structures in combination with new technology regulation are not sufficient to deal with the social and ethical consequences of algorithmic technologies discussed above (Van Brakel, 2021). To understand and assess the potential social and ethical concerns of these systems it is not only important to

focus on the technology itself but also the social structures they are embedded in as the issues and risks are not only the result of the technology itself but can, for instance, already be inherent in the social practices themselves or concerns that are raised by the intervention as the result of the use of technology. For instance, in the above-mentioned Top400 list the intervention that follows as the result of being put on the list by an algorithm is what is causing the most harm and not only for the young people themselves but also for their relatives such as their mothers (Van Brakel and Govaerts, 2024).

Conclusion

AI will have a significant impact on crime and crime control by automating current practices but also by shifting work practices. These developments open new and important research avenues within criminology and more specifically digital criminology. Digital criminological research can contribute to exploring and setting the socio-technical boundaries within which AI can be implemented or not implemented in crime control. Moreover, university degrees in criminology should take up digital criminology as part of the programmes so that future criminal justice practitioners and policymakers have the necessary expertise to be able to use and assess the use of AI critically. Finally, criminologists, have a unique position and expertise to make a significant contribution to regulatory and policy discussions with regards to crime control and criminal justice and more specifically when implementing AI in policing, prisons, and in the context of crime prevention programs and criminal psychology, but also when AI is being used to commit crime. This implies integrating knowledge from qualitative and ethnographic but also critical criminological research into discussions and policymaking on AI.

Suggested reading

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing & Society*, 28(7), 806 – 822.

Schuilenburg, M., & Peeters, R. (eds.) (2021). *The Algorithmic Society: Technology, Power and Knowledge.* London: Routledge.

Van Brakel, R. & Govaerts, L. (2024). Exploring the impact of algorithmic policing on social justice: Developing a framework for rhizomatic harm in the pre-crime society. *Theoretical Criminology, OnlineFirst.* Available at: https://doi.org/10.1177/13624806241246267.

References

Babuta, A., & Oswald, M. (2021). Machine learning predictive algorithms and the policing of future crimes: Governance and oversight. In J. L. M. McDaniel & K. Pease (eds.), *Policing and Artificial Intelligence* (pp. 214 – 226). London: Routledge.

Bender, E., Gebru, T. McMillan, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?, *FAccT '21*, March 3 – 10, 2021, Virtual Event, Canada.

Buolamwini, J., & Gebru. T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1 – 15.

European Commission. (2021). COM (2021) 206 final Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts.

European Parliament. (2021). European Parliament resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html.

European Parliament (2024) Regulation (EU) 2024/1689 Artificial Intelligence Act, *OJ L, 2024/1689, 12.7.2024*, 1 – 144.

Europol. (2021). *Malicious Uses and Abuses of Artificial Intelligence Europol's European Cybercrime Centre (EC3).* https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf.

Fussey, P., & Murray, D. (2019). *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology.* (Human Rights and Big Data Project, University of Essex). Available at: https://48ba3m4eh2bf2sksp43rq8kkwpengine (accessed 10 December 2023).

Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender. *Social Sciences*, 12(8), 435. Available at: https://doi.org/10.3390/socsci12080435 (accessed 10 December 2023).

House of Lords Communications and Digital Committee. (2024). *Large Language Models and Generative AI,* 1st Report of Session 2023 – 24. Available at: https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/5402.htm (accessed 10 December 2023).

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.

Kaun, A., & Stiernstedt, F. (2020). Doing time, the smart way? Temporalities of the smart prison. *New Media & Society*, 22(9), 1580 – 1599.

Liu, H.-Y., Maas, M., Danaher, J., Scarcella, L., Lexer, M., & Van Rompuy, L. (2020). Artificial intelligence and legal disruption: A new model for analysis. *Law, Innovation and Technology*, 12(2), 205 – 258.

Lynch, S. (2017). Why AI is the new electricity? *Insights by Stanford Business*, 11 March 2017. Available at: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity (accessed 10 October 2023).

National Academies of Sciences, Engineering, and Medicine. (2024). *Facial Recognition Technology. Current Capabilities, Future Prospects, and Governance.* Washington, DC: The National Academies Press. Available at: https://doi.org/10.17226/27397.

Oliveira, A. L., & Figuerido, M. A. T. (2024). Artificial Intelligence: Historical context and state of the art. In H. Sousa Atunes, P. M. Freitas, A. L. Oliveira, C. M. Pareira, E. Vaz de Sequeira, & L. Barreto Xavier (eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (pp. 3 – 24). Dordrecht: Springer.

Oswald, M., Grace, J. Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'Experimental' proportionality. *Information and Communicaiton Technologies Law*, 27(2), 223 – 250.

Phelps, S. (2023). Beyoncé's Cardiff gig crowd was scanned for paedophiles. *BBC News*, 10 November 2023. Available at: https://www.bbc.com/news/uk-wales-67360696 (accessed 10 November 2023).

Powell, A., Stratton G., & Cameron, R. (2017). Crime and justice in digital society: Towards a 'digital criminology'? *International Journal for Crime, Justice and Social Democracy*, 6(2), 17 – 33.

Puolakka, P., & Van De Steene, S. (2021). Artificial Intelligence in prisons in 2030. An exploration of the future of AI in prisons. *Advancing Corrections Journal*, 11, 128 – 138.

Sandhu, A., & Fussey, P. (2020). The 'uberization of policing'? How police negotiate and operationalise predictive policing technology. *Policing & Society*, 31(1), 66 – 81.

Turing, A. (1950). Computing machinery and intelligence, *Mind*, 59(236), 433 – 460.

UNESCO (2023) *Technology-Facilitated Gender-Based Violence in an Era of Generative AI.* Available at https://unesdoc.unesco.org/ark:/48223/pf0000387483/PDF/387483eng.pdf.multi (accessed 12 November 2023).

Van Bekkum, M., & Borgesius-Zuiderveen, F. (2021). Digital welfare fraud detection and the Dutch *SyRI* judgment. *European Journal of Social Security*, 23(4), 323 – 340.

Van Brakel, R. (2016a) Pre-emptive Big Data surveillance and its (dis)empowering consequences: The case of predictive policing. In B. van der Sloot, D. Broeders, & E. Schrijvers (eds.), *Exploring the Boundaries of Big Data* (pp. 117 – 141). Amsterdam: Amsterdam University Press.

Van Brakel, R. (2016b). The rise of pre-emptive surveillance of children in England: Unintended social and ethical consequences. In E. Taylor & T. Rooney (eds.), *Surveillance Futures: Social and Ethical Implications of New Technologies on Children and Young People* (pp. 190 – 203). London: Routledge.

Van Brakel, R. (2021). How to watch the watchers? Democratic oversight of algorithmic police surveillance in Belgium. *Surveillance & Society*, 19(2), 228 – 240.

Van Brakel, R. (2023). Artificiële Intelligentie en veiligheid: Kansen en uitdagingen. *Politie en Recht,* 4, 151 – 162.

Van Brakel, R. & Govaerts, L. (2024). Exploring the impact of algorithmic policing on social justice: Developing a framework for rhizomatic harm in the pre-crime society. *Theoretical Criminology*, *OnlineFirst.* Available at: https://doi.org/10.1177/13624806241246267.

Van Schendel, S. (2019). The challenges of risk profiling used by law enforcement: Examining the cases of COMPAS and SyRI. In L. Reins (ed.), *Regulating New Technologies in Uncertain Times* (pp. 225 – 240). Dordrecht: Springer.

Wang, A., Kapoor, S., Barocas, S., & Narayanan, A. (2023). Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. *ACM Journal on Responsible Computing*, 1(1), 1 – 45. Available at: https://dl.acm.org/doi/10.1145/3636509 (accessed 2 February 2024).

Wilson, D., & Norris, C. (eds.). (2006). *Surveillance, Crime and Social Control.* London: Routledge.

Monique Mann

10 Automation

Abstract: This chapter explores the history and meaning of automation and its relevance in criminal justice. It is argued that there is a need to think critically about what 'automation' means and attend to the social context/s, socio-political systems, actors, and networks within them, as well as specific applications of 'automated' technologies to understand drivers and consequences at the intersection of science, technology, and society.

Keywords: automation, criminal justice, bias, discrimination

Introduction

The origins of automation stem from factory work and manufacturing with the objective of increasing efficiency, which over time has expanded to other areas such as health care, transportation, agriculture, construction, government administration, military, security, and criminal justice (Hitomi, 1994). There are various applications of automation in criminal justice. It is used, for example, to detect intrusions in computer systems, to predict recidivism, and to impose sentences or financial penalties. The use of automation in criminal justice is the focus of this contribution.

The term 'automaton' was coined by D. S. Harder while working for the General Motors Corporation in 1936, who later went on to establish the Automation Department at the Ford Motor Company (Hitomi, 1994). Therefore, automation has its origins in Fordist modes of production. At the time, Harder defined automation "as the transfer of work parts between the machines in a production process *without human operation*" (Hitomi, 1994: 122, emphasis added). However, this view of automation is deceptively simple and reductionistic as it does not consider wider socio-technical *assemblages* that shape 'automated' systems and processes (for an overview of the philosophical concept of 'assemblage' see Marcus and Saka, 2006).

'Automated' systems and processes may appear to operate independently of humans, but ultimately, they do not. 'Automation' is more complex than an absence of human influence, as it is constructed, enabled, and enacted by socio-political and human dimensions. These include decisions to collect information (or not), the analysis of information in certain ways for some purposes (and not others), the programming of algorithms to perform one function (over another), and they extend to how 'automated' outputs guide and influence human behavior (or not) (see Algorithm by Leese). Automation is always part of a wider *assemblage*, and as will be argued in this chapter, this concept is helpful to both understand and problematize 'automation' in criminal justice.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-011

Notwithstanding this critique, which is expanded on further below, historically, there are three main concepts of automation, as follows (developed by Diebold, Bright, and Druker, as cited in Hitomi, 1994: 122):

> 'Automation' may be considered the abbreviation of 'automatisation' or 'automatic operation'; alternatively automation is a combination of the Greek 'automotos' (meaning self-moving) and Latin –'ion' (meaning a state). 'Mechanization' is the replacement of human physical labour by machines, but the control of this machine operation is effected by human operators. However, 'automation' also replaces this control action by machines; that is 'automation' means the replacement of both human physical and mental activities by machines.

There is an important distinction, however, between automation and robotics (see Robots by Sandvik and Lintvedt). The IEEE Robotics and Automation Society Field of Interest Statement defines robotics as focusing on "systems incorporating sensors and actuators that operate autonomously or semi-autonomously in cooperation with humans" and defines automation (or research on it) with an emphasis on "*efficiency*, *productivity*, *quality*, and *reliability*, focusing on *systems that operate autonomously*, often in *structured environments* over *extended periods*, and on the explicit *structure of such environments*." (Goldberg, 2012: 1, emphasis added). It is necessary, therefore, to consider the structure of environments which can be extended to include the social and political contexts in which 'automation' is both designed and deployed.

Relevance of automation in criminal justice

Automation in criminal justice has immense intellectual and criminological relevance as there has been an exponential increase in the collection, aggregation, and *automated* analysis of information and *automated* decision-making within the criminal justice system. This includes applications in digital forensics (Jarett and Choo, 2021), policing including predictive, pre-emptive, and intelligence-led approaches (Bennett Moses and Chan, 2014; 2016; Egbert and Mann, 2021), courts and sentencing, especially risk assessment (Završnik, 2019; 2020), and within prisons and carceral contexts (McKay, 2022). 'Big data' inform automation and automated decision-making, which has been termed "stochastic governance" (Sanders and Sheptycki, 2017). It has been argued that new forms of 'automatic justice' or 'automated justice' are challenging the traditional model of criminal justice and the discipline of criminology (Bowling et al., 2008; Chan and Bennett Moses, 2016; Marks et al., 2017; Brownsword and Harel, 2019). This is because divisions between surveillance, policing, adjudication, and punishment are being eroded with automation collapsing these processes and removing human actors, at least from obvious involvement (Marks et al., 2017; Završnik, 2019).

Automation may have the potential to improve the efficiency and accuracy of decision-making under certain circumstances or for some applications, but it can be used in ways that are harmful (Bennett Moses and Chan, 2014). This may include biases in datasets or the parameters of algorithms that underpin and facilitate automation tar-

geting 'risky' individuals (Friedman and Nissenbaum, 1996), or automated technologies deployed against the most marginalized. These include populations that are more likely to be subject to disciplinary surveillance and control (Monahan, 2017) employed to 'sort' and subject them to differential (or discriminatory) treatment (Lyon, 2003). Automation in criminal justice is problematic as it involves the increased monitoring of people released from prison based on actuarial risk assessments (Harcourt, 2005). These processes are never neutral, and bias or errors may be difficult to contest with barriers to transparency and a false veil of objectivity provided by computerization (Friedman and Nissenbaum, 1996).

Conceptual complications of 'automation'

There are different conceptual perspectives that can help understand, and complicate, 'automation' in criminal justice. Automated processes require input data and through analyzing information, automated decisions can be made, or automated actions completed. Therefore, the starting point that enables automation, is the collection of information via forms of surveillance (see Lyon: Surveillance), that may also be automated (Andrejevic, 2019). Significantly, surveillance practices and datasets converge into a "surveillant assemblage" (Haggerty and Ericson, 2000). Rather than involving independent datasets, discreet technologies or surveillance practices, Haggerty and Ericson (2000: 610) argue that "surveillance is driven by the desire to bring systems together, to combine practices and technologies and integrate them into a larger whole." They also draw attention to the fact that surveillance is often directed on or towards the human body (for example, a face via facial recognition technology). According to Haggerty and Ericson (2000: 612–613) "the surveillant assemblage relies on machines to make and record discrete observations" by standardizing "the capture of flesh/information flows of the human body" and creating "data doubles."

Automated decisions upon "data doubles" are occurring in many social institutions, not only criminal justice, but also, for example, welfare administration. Take, for instance, my own case study of the 'RoboDebt' automated debt raising scandal (Mann, 2020). This involved *automated* income matching of data about individuals held by and shared between various government agencies to *automatically* identify overpayment of benefits, and the *automated* issuance of debts *automatically* sent to welfare recipients. However, the program was driven by politics, rather than automated technologies (Mann, 2020).

Haggerty and Ericson's (2000) "surveillant assemblage" and this example of automated debts connects to O'Malley's (2010) concepts of "simulated justice" and "telemetric policing," which take the "surveillant assemblage" further by theorizing shifts in the administration of 'justice' to the "monetisation of justice" (O'Malley, 2010: 795). O'Malley (2010) uses the example of *automatic* traffic fines, one that will be familiar to many readers: an offense is *automatically* detected by a camera, an individual (or its "data double") is *automatically* identified by a number plate, connected to information stor-

ed in a roads traffic authority database (i. e., the "surveillant assemblage"), all of which leads to a fine being *automatically* posted. This occurs with little apparent involvement of criminal justice agents such as the police or judges. O'Malley (2010: 797) explains that the "nexus with money" is a central aspect of "simulated justice" as it enables financial penalties to be *automated.* Yet, humans are involved in conceptualizing, constructing, and authorizing the operation of such 'automated' systems, which are influenced by wider social and political dynamics such as the desire to improve road safety or raise revenue. So, are they truly 'automated'?

'Automated,' 'assisted,' or an 'assemblage'?

Facial recognition technology (FRT) is one technology that is often considered to be 'automated' (Mann and Smith, 2017). It is used here as an example to highlight the importance of the contexts and applications of technologies in criminal justice, and to question whether they are indeed 'automated.'

The main applications of FRT, at present, are one-to-one matching of biometric templates to automatically verify the identity of a person. This occurs at a border with access being granted if the biometric template of the individual physically standing at the smart gate matches the biometric template stored in their passport, and *provided that the individual, their passport and their 'data double' are deemed worthy of access.* This is a process of social sorting that reflects geo/political decisions about borders and citizenship (Lyon, 2003).

The second main application of FRT involves one-to-many searching. This could involve the CCTV footage, images from the internet, or those stored in other databases to identify someone unknown (Mann and Smith, 2017). The concepts of the 'surveillant assemblage,' 'simulated justice,' and 'telemetric policing' are relevant. This is because information drawn from the 'surveillant assemblage' is used to automatically identify an individual who may be deemed 'risky' (an assessment made from information held in databases and analyzed in automated ways; see Databases by Bellanova), or subject to an automated fine due to behavior remotely and automatically identified through surveillance.

It has been argued that facial recognition technology, when used for one-to-many searching, is not *automated per se,* as technologies and humans converge in hybrid ways and operate as part of a wider assemblage (Fussey et al., 2021). Fussey et al. (2021) argued that '*assisted*' FRT is a more appropriate term, given the ways in which FRT is used by police, for example by deciding about whether (or not) to stop someone that the FRT algorithm identifies. Police make discretionary decisions following identification by a FRT algorithm. Hence, one would rather speak of a hybrid interaction between police and the FRT algorithm (Fussey et al., 2021).

This is related to Egbert and Mann's (2021) critique of predictive policing technologies and their analysis of the ways in which police act upon the outputs of them (see Policing by Wilson; Prediction by Ķīlis, Gundhus, and Galis). For example, police may

behave differently in areas that are predicted to be either high or low risk (Egbert and Mann, 2021). What is more, police may be more attentive in areas predicted to be 'high risk.' Furthermore, there are a range of factors that shape the inputs of predictive technologies, including the data that informs the algorithms, which are collected under certain social and political conditions, and are not neutral either. An example of such data are the types of crimes reported to or detected by police (Egbert and Mann, 2021).

The point of discussing the distinction between '*automated*' and '*assisted*' using the examples of FRT and predictive policing is to critique and problematize reductionistic conceptual binaries between 'automated' or otherwise. *All* technologies and their applications are mediated by a range of complex social, political, bureaucratic/institutional, and human factors within a wider assemblage. This has consequences for bias and discrimination (and not *only* in a technical sense), for example when 'automated' technologies in criminal justice tend to be deployed towards the most marginalized and those that already disproportionately have contact with the criminal justice system due to historical, social, and political reasons. Automated technologies do not exist in a social, political, or institutional vacuum.

Issues and concerns with automation in criminal justice

There are international attempts to regulate automation and automated decision-making through, for example, Article 22 of the European Union's General Data Protection Regulation (GDPR) that grants a right not to be subject to automated decisions and introduces a so-called 'right to explanation' (on limits to data protection and a 'right to explanation', see Wachter et al., 2017; Mann and Matzner, 2019). However, issues and concerns, especially given the potential human rights impacts, remain (Završnik, 2020). Examples of rights that may be impacted by automation include non-discrimination, especially when the logic of algorithms is to discriminate (Mann and Matzner, 2019).

Moreover, automated technologies operate at scale (one scale is, for example, population level in the case of traffic fines). There is a risk that due process safeguards may be undercut. An example includes the presumption of innocence and punishment for crimes committed predicted or pre-empted based on risk rather than committed with criminal conviction (see Garrett, 2013). A further example includes the right to a fair trial and being able to hear and understand the evidence, which may prove challenging with propriety algorithms. There are also a range of concerns such as biases that continue historical social trajectories and discriminate against women and People of Colour (Buolamwini and Gebru, 2018). It is often argued that it is necessary to have a 'human in the loop' to safeguard from the negative consequences of automation (see for example Lettieri et al., 2023). However, this may introduce other issues as 'assisted' or human decision-making are not free from bias either, as I have argued above and

elsewhere (Egbert and Mann, 2021, see also Enarsson et al., 2021). This means that both human and assisted decision-making can be biased, with the solution to address bias. Therefore, this requires socio-political and structural solutions rather than only technological ones. Hence, the interaction/s between automation by computers (and datasets, analytic processes, and algorithms), socio-political systems, and humans as a complex assemblage demands greater study, theorizing, and empirical analysis.

Conclusion and future directions

Technology is developing rapidly, and the algorithms at the base of automation and automated decisions are becoming more complex (for example, 'Artificial Intelligence' or 'AI'). The number of sensors is increasing (e. g., at home, in the car, or on/in the body), a trend that will continue feeding the 'surveillant assemblage.' There will be more 'automated' 'telemetric policing' and 'simulated justice' opportunities. Automation will also operate in combination with and/or within other emerging policing technologies (such as forms of *robotic* policing). There are also highly controversial applications of automation that have been advocated for and may develop over time. For example, in the field of imprisonment this would involve automation for remote monitoring via surveillance systems and electronic monitoring devices such as ankle bracelets. These monitoring devices can include, for example, automated conducted energy devices and functionalities to 'automatically' immobilize (rather than fine) (Bagaric et al., 2018). Such projects demonstrate the importance of thinking critically about the applications for automation, and the impacts for individuals, their rights, and the principles of criminal justice.

Main takeaways

- The collection, assemblage, and analysis of information and associated 'automated' decisions or actions in criminal justice contexts is increasing. Concepts such as assemblage/s can help understand, critique, and problematize 'automation.'
- There is a need to think critically about what 'automation' means and attend to the social context/s, socio-political systems, actors, and networks within them, and specific applications of 'automated' technologies to understand drivers and consequences at the intersection of science, technology, and society.
- Automation presents risks of bias, discrimination, errors, and introduces challenges regarding transparency, fairness, justice, and recourse for those subject to it.
- There is an immediate need for strong and enforceable legal and regulatory frameworks to protect human rights and ensure due process. While some international models emerge (such as Article 22 of the GDPR or the more recent European Artificial Intelligence Act, but those have limits too), legal and regulatory issues still need discussion (see, for example, Wachter et al., 2017; Mann and Matzner, 2019).

Suggested reading

Andrejevic, M. (2019). Automating surveillance. *Surveillance and Society*, 17(1/2), 7 – 13.
Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York City: St. Martin's Press.
Marks, A., Bowling, B., & Keenan, C. (2017). Automatic justice? Technology, crime and social control. In R. Brownsword, E. Scotford, & K. Yeung (eds.) *The Oxford Handbook of Law, Regulation and Technology* (pp. 705 – 730). Oxford: OUP.

References

Andrejevic, M. (2019). Automating surveillance. *Surveillance and Society*, 17(1/2), 7 – 13.
Bagaric, M., Hunter, D., & Wolf, G. (2018). Technological incarceration and the end of the prison crisis. *The Journal of Criminal Law and Criminology*, 108(1), 73 – 135.
Bennett Moses, L., & Chan, J. (2014). Using big data for legal and law enforcement decisions: Testing the new tools. *UNSW Law Journal*, 37(2), 643 – 678.
Bennett Moses, L., & Chan, J. (2016). Algorithmic prediction in policing: Assumptions, evaluation and accountability. *Policing and Society*, 8(7), 806 – 822.
Bowling, B., Marks, A., & Murphy, C. (2008). Crime control technologies: Towards an analytical framework and research agenda. In R. Brownsword & K. Yeung (eds.), *Regulating Technologies* (pp. 51 – 78). Oxford: Hart Publishing.
Brownsword, R., & Harel, A. (2019). Law, liberty and technology: Criminal justice in the context of smart machines. *International Journal of Law in Context*, 15(2), 107 – 125.
Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 81, 77 – 79.
Chan, J., & Bennett Moses, L. (2016). Is big data challenging criminology? *Theoretical Criminology*, 20(1), 21 – 39.
Egbert, S., & Mann, M. (2021). Discrimination in predictive policing: The dangerous myth of impartiality and the need for STS-Analysis. In A. Završnik & V. Badalic (eds.), *Automating Crime Prevention, Surveillance, and Military Operations* (pp. 25 – 46). Cham: Springer.
Enarsson, T., Enqvist, L., & Naarttijarvi, M (2021). Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law*, 1, 123 – 153.
Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330 – 347.
Fussey, P., Davies, B., & Innes, M. (2021). 'Assisted' facial recognition and the reinvention of suspicion and discretion in digital policing. *British Journal of Criminology*, 61, 325 – 344.
Garrett, B. L. (2013). Big data and due process. *Cornell Law Review Online*, 99, 207 – 217.
Goldberg, K. (2012). What is automation? *IEEE Transactions on Automation Science and Engineering*, 9(1), 1 – 2.
Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *British Journal of Sociology*, 51(4), 605 – 622.
Harcourt, B. (2005). Against prediction: Sentencing, policing, and punishing in an actuarial age. *University of Chicago Public Law & Legal Theory Working Paper, No. 94.*
Hitomi, K. (1994). Automation – its concept and a short history. *Technovation*, 14(2), 121 – 128.

Jarett, A., & Choo, K. R. (2021). The impact of automation and artificial intelligence on digital forensics. *WIREs Forensic Science*, 3(6), 1 – 17.

Lettieri, N., Guarino, A., Zaccagnino, R., & Malandrino, D. (2023). Keeping judges in the loop: A human–machine collaboration strategy against the blind spots of AI in criminal justice. *Soft Computing*, 27, 11275 – 11293.

Lyon, D. (2003). Surveillance as social sorting: Computer codes and mobile bodies. In D. Lyon (ed.), *Surveillance as SOCIAL SORTING: Privacy, Risk and Digital Discrimination* (pp. 13 – 30). London and New York: Routledge.

Mann, M. (2020). Technological politics of automated welfare surveillance. *Global Perspectives*, 1(1), 1 – 12.

Mann, M., & Matzner, T. (2019). Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data and Society*, 6(2), 1 – 11.

Mann, M., & Smith, M. (2017). Automated facial recognition technology: Recent developments and strengthening oversight. *UNSW Law Journal*, 40(1), 121 – 145.

Marcus, G. E., & Saka, E. (2006). Assemblage. *Theory, Culture and Society*, 23(2 – 3), 101 – 109.

Marks, A., Bowling, B., & Keenan, C. (2017). Automatic justice? Technology, crime and social control. In R. Brownsword, E. Scotford & K. Yeung (eds.), *The Oxford Handbook of Law, Regulation and Technology* (pp. 705 – 730). Oxford: OUP.

McKay, C. (2022). The carceral automation: Digital prisons and technologies of detention. *International Journal of Crime Justice and Social Democracy*, 11(1), 100 – 119.

Monahan, T. (2017). Regulating belonging: Surveillance, inequality, and the cultural production of abjection. *Journal of Cultural Economy*, 10, 191 – 206.

O'Malley, P. (2010). Simulated justice: Risk, money and telemetric policing. *British Journal of Criminology*, 50, 795 – 807.

Sanders, C., & Sheptycki, J. (2017). Policing, crime and 'big data'; towards a critique of the moral economy of stochastic governance. *Crime, Law and Social Change*, 68, 1 – 15.

Wachter, S., Mittelstadt, B., and L. Floridi. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76 – 99.

Završnik, A. (2019). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 18(5), 623 – 773.

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*, 20, 567 – 583.

Marion Oswald and Angela Paul

11 Bias

Abstract: This chapter presents examples of biases that can occur in relation to the use of data-driven and algorithmic tools within criminal justice, and explores the different definitions of 'bias' in criminology, computer science, and law. We highlight scholarly discourse on analysis, as well as risk and mitigation from the perspectives of the above disciplines, and offer a new taxonomy to aid researchers. The chapter advocates for an interdisciplinary strategy for understanding and regulating data-driven approaches within criminal justice.

Keywords: bias, policing, algorithms, law, discrimination

The origin of the word 'bias,' although contested, is thought to have derived from the Old French word 'biais,' which means slant, slope, or against the grain (CEBM, 2018). The word may have also entered the English language through the game of bowls, which is played with 'biased' balls that travel obliquely because of the difference in weight on one side of the balls. The term 'bias' is still used in everyday language to refer typically to a psychological inclination—sometimes a reprehensible one—towards or against a certain opinion. However, the term takes on alternative or additional nuances when employed by law, computer science, and criminology/surveillance studies in the context of the use of algorithmic tools in criminal justice (see Algorithm by Leese). The daily operations of law enforcement and the legal system have been revolutionized by the incorporation of machine learning techniques, sometimes replacing human expertise with algorithmic evaluations (see Artifical Intelligence by Van Brakel).

This chapter explores definitions and understandings of the term 'bias' common to the disciplines of law, computer science, and criminology, which are crucial to understanding the significance of bias in digital criminology. By doing so, we highlight similarities and differences in such definitions and understandings. The chapter presents selected examples of biases that can occur in relation to the use of data-driven and algorithmic tools within criminal justice, highlighting scholarly discourse on analysis, as well as risk and mitigation from the perspectives of the above disciplines. The term bias is frequently used to refer to unlawful discrimination based on protected characteristics, such as race, which can commonly occur in the criminal justice system (Richardson et al., 2019). We offer a taxonomy of bias to help researchers looking at bias in these contexts. Although out of scope for this chapter, we note the overlap between issues of bias and other important rights, including those relating to privacy and data

Disclaimer: Angela Paul's PhD studies were supported by Northumbria University.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-012

protection, and the issue of collection and use of sensitive demographic data for the purposes of mitigating bias (CDEI, 2023).

Understandings of bias in the context of algorithmic criminal justice

It is perhaps in the legal context that we come closest to the term 'bias' reflecting common usage. This is because bias is a concept that is relevant to the governing of state power, and to the issue of unlawful discrimination. For example, in English common law, judicial review is the procedure whereby the courts supervise the lawfulness of the exercise of power by public bodies. A key ground of review (and element of natural justice) is the 'rule against bias.' That is, to ensure a decision-making process is fair, no proven or real danger of bias should be present, assessed from the point of view of a fair-minded and informed observer. What does bias mean in this context, however? Law recognizes that bias in human decision-making represents a mental inclination, even one unknown to the decision-maker, as per Lord Goff:

> bias is an insidious thing that, even though a person may in good faith believe he was acting impartially, his mind may unconsciously be affected by bias (*R v Gough*, 1993).

The issue of bias is linked closely to the need to ensure the fairness of any tribunal or decision-making process (Article 6, ECHR, emphasizes the right to a fair trial by an independent and impartial tribunal), as stated by Lord Hope:

> The word 'bias' is used as a convenient shorthand. But it would be a mistake to approach it in this context as if its only meaning were pejorative. The essence of it is captured in the Convention concept of impartiality. An interest in the outcome of the case or an indication of prejudice against a party to the case or his associates will, of course, be a ground for concluding that there was a real possibility that the tribunal or one of its members was biased … but the concept is wider than that. It includes an inclination or pre-disposition to decide the issue only one way, whatever the strength of the contrary argument. (*Davidson v Scottish Ministers*, 2004)

Although bias in this context is concerned with bias (or reasonable suspicion of bias) of the individual decision-maker or tribunal, the insertion of an algorithm into the process may also create or exacerbate that bias. As noted by Oswald et al., "it could be said that a risk-averse algorithm, which we know over-estimates risk in order to maximise public protection but which generates a degree of 'false positives' of high-risk results to do this, might actually be creating a biased process (or tribunal of sorts)" (Oswald et al., 2018: 241.). For example, a facial recognition tool might have a low face-match threshold in order to minimize the risk of missing a wanted individual, but this then results in a higher risk of false positives occurring (see Facial Recognition by Fussey). Similarly, Cobbe points out that due to less favorable treatment or disadvantage reflected in training data, "ADM [automated decision-making] systems may be prone to making de-

cisions which are systematically skewed in some way, rather than acting impartially." Objective judgment could be prevented by "the presence of an internal model which does not produce fair and consistent outputs (for example, a system could, without any intention to do so on the part of the public body, treat those from certain socio-economic backgrounds less favourably than others)" (Cobbe, 2019).

The biases which are present in automated systems are referred to as 'statistical biases,' which are caused by using data that includes systematic errors which, in turn, skew results. These errors can arise from the algorithm itself and/or from faulty, incomplete, irrelevant, excluded, or biased data sets, thus creating consistent errors in a model, the error being the difference between the ground truth and the average model prediction (i. e., between the reality on the ground and what the model is telling you). Bias is also often used as a shorthand, for unlawful discrimination on the grounds of certain protected characteristics such as sex, gender, age, sexual orientation, disability, and race. These two understandings of bias are closely connected as we explore below.

Machine learning tools have been incorporated into the daily work and processes of police and the judicial system, for instance, to replace the expert opinions of psychiatrists or probation officers, which were used in the past to judge the risk of criminality (see Policing by Wilson). The main justification given for using risk assessment instruments can be to "reduce the noise inherent to human decision making … for example, some judges predict recidivism better than others" (Goel et al., 2018). However, the use of data analytics to make law enforcement decisions or predictions may mean that disparities become encoded into the datasets, thus feeding back into the system. In the context of the enforcement of marijuana violations in the US, Butcher et al. argue that "predictions made by these tools may reflect or even exacerbate past racially disparate enforcement" (Butcher et al., 2022: 137). Recidivism is sometimes predicted using data inputs that include certain personal characteristics, for instance, an individual's neighborhood or socioeconomic status. Although race may not be an explicit predictor, algorithms can in fact "contain an imperfect proxy for race" or other protected characteristics (Davies and Douglas, 2020), for instance the risk of using arrest in tools that aim to predict risk of re-offending (Fogliato et al., 2021).

Discussions of bias, in terms of unlawful discrimination that can result from it, are extremely relevant in the context of automated criminal justice processes. A type of statistical bias is 'sampling bias,' which can occur when systemic discrimination is reflected in training data (Borgesius, 2018). Systems built on data from periods of flawed, racially biased, and potentially unlawful practices ('dirty data') could result in flawed predictions and harmful feedback loops (Richardson et al., 2019). Birhane too calls for understanding of historical injustices and power asymmetries embedded within algorithmic systems (Birhane, 2022). In addition, a study which surveyed legal professionals in the UK revealed that racial bias plays a significant role in the judicial system, and 55.6 % of the legal professionals said that they have "witnessed one or more judges act with racial bias in their treatment of defendants" (Monteith et al., 2022). Furthermore, the Casey Review into the UK's Metropolitan Police found that:

> Black Londoners in particular remain over-policed. They are more likely to be stopped and searched, handcuffed, batoned and Tasered, are over-represented in many serious crimes, and when they are victims of crime, they are less satisfied with the service they receive than other Londoners. (Baroness Casey, 2023: 17)

These revelations around 'dirty data,' and the witting or unwitting institutional biases that are already embedded in the criminal justice systems, may correlate to the biases that can be present in the outputs of digital analytical processes.

There is of course a high risk that the use of a biased algorithm could result in unlawful direct or indirect discrimination on the grounds of protected characteristics, or breach of other equality duties, contrary to legislation such as the UK's Equality Act 2010. We refer to this concept as 'statistics-based discrimination,' where actions or decisions taken based on or guided by the biased output contribute to prohibited conduct. We would need however to consider the link between the biased algorithm and the unlawful discrimination (or positive duties to have due regard to eliminating such discrimination: s149(1) Equality Act 2010). For example, Allen and Masters give an example of AI-driven online advertising resulting in women being shown a job advert less frequently than men. This is likely to be unjustified indirect discrimination contrary to s19 of the Equality Act (Allen and Masters, 2021). Another example can be seen from the famous analysis by journalists at ProPublica who concluded that the risk scores generated by a system called Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), used to inform sentencing decisions, were nearly twice as likely to predict falsely that black defendants would commit crime in the future when compared to white defendants (Angwin et al., 2016). Allen and Masters describe this as creating "a new stereotype" (Allen and Masters, 2021: 44).

Furthermore, an evaluation undertaken by the UK's National Physical Laboratory of the facial recognition system used by the Metropolitan Police concluded that at a face-match threshold of 0.6, the system operated equitably between gender and ethnicity. However, if a lower threshold was used, then ethnic disparities increased. At the 0.56 setting, of the 33 people falsely identified in the trial, 22 were black, 8 were Asian, and 3 were white, 18 being in the 21–30 age group (Mansfield, 2023: 7). The bias in the system revealed at this setting would have direct implications for the justifiability, reliability, and lawfulness of decisions informed by the probabilistic output, including stop and search, and even arrest. Cases from the US suggest that despite guidance being issued that a 'match' can never be a sole ground for arrest, officers tend to rely on the outputs of facial recognition and wrongful, biased arrests have occurred (Magnet, 2011: 150; Johnson, 2022; Bhuiyan, 2023).

We have focused so far on data bias connected to discrimination on the grounds of protected characteristics. However, bias as it relates to the deployment of algorithms within criminal justice may impact wider decision-making. Algorithmic decision-making and data analytics may create new categories or groups on the basis of criteria that are not necessarily linked to protected characteristics, but may still cause unfairness, for example due to social inequalities (Gerards and Borgesius, 2022). Wachter, Mittel-

stadt, and Russell point to certain fairness metrics, such as equalized odds, classification parity, or false positive error rate equality, that may in fact preserve bias rather than reducing it, because these methods do nothing to address underlying causes of inequality (Wachter et al., 2021).

This concern could arguably be illustrated by the assessment of equitability carried out on the facial recognition system used by the Metropolitan Police as mentioned above (Mansfield, 2023: 7) which defines equitability between demographics as requiring that, "in the operational setting, the outcomes for the subjects (i. e., recognition rates and false alert rates) should be broadly equivalent for demographics considered." False alerts may still occur however which may raise broader fairness issues. Computational methods of determining fairness or bias could be insufficient to incorporate the interpretative and contextual legal tests of proportionality and legitimacy (Sanchez-Monedero et al., 2020). In respect of probabilistic classifications (i. e., a statistical determination that someone/thing might meet a group's characteristics), it may be impossible to satisfy the conditions of competing notions of fairness simultaneously (Kleinberg et al., 2017).

Babuta and Oswald argue that focusing on 'data' bias may distract attention away from the issue of whether algorithmic techniques are appropriate at all for particular criminal justice contexts (Babuta and Oswald, 2019). Bias may occur at all stages of the project lifecycle, from bias in favor of data-driven solutions, moving to bias in training data and misleading accuracy rates. Issues of 'model fit' may occur if an algorithm is trained and validated only on a limited sample, and then applied to a more diverse dataset, for example if an algorithm is trained only on older males, and then applied to a dataset that includes females and younger adults. Automation bias may be an issue, a form of cognitive bias in which individuals are partial towards the decisions of automated systems, leading to the individual deferring to automated decisions over human judgment (Citron, 2008; Rieke and Bogen, 2018: 9). Oversight itself could be biased if limited to data science expertise or where there is a lack of independence (see figure 1, Risk of bias in implementation and oversight of police algorithms in Babuta and Oswald, 2019).

Furthermore, Kaufmann argues that the concept of bias itself could be misleading or superfluous: "In understanding information as relational, context-specific and lively, bias becomes a superfluous concept, because it is everywhere, in every dataset. The only way to engage with 'bias' is then to identify and reflect about the specificities of information and how to engage with them" (Kaufmann, 2023: 16).

Conclusions and lessons for research

The deployment of digital, algorithmic, and AI technologies within criminal justice brings with it a risk of bias. Yet researching such 'bias' also brings the risk of misunderstandings between disciplines if the term and the type(s) of bias in play or to be investigated are not clearly defined from the outset. We set out below a classification

of the types of bias explored in this chapter which can be deployed by researchers in this field at the outset of their projects in order to minimize the likelihood of working at cross-purposes:

Table 1: Bias classification

***Type* of bias**	***Elements* and *location* of this bias**
Cognitive or decision-making	Inclination to decide one way due to partiality—human decision-maker or systematically skewed algorithm. Legal bias—The rule against bias in English common law through judicial review i. e. to ensure a tribunal or decision-making process is impartial and fair. Automation bias—human decision-maker deferring to the algorithmic output.
Statistical	The overarching term for statistically driven decision-making that can be based on faulty, incomplete, irrelevant, excluded. or biased data sets. Sampling bias—systemic discrimination is reflected in automated decision-making due to biased training data, incomplete data or proxy variables. Model fit—where an algorithm is trained and validated only on a limited sample, and then applied to a more diverse or different dataset. Statistics-based discrimination—where actions or decisions based on, guided by, or linked to the biased output contribute to prohibited/unlawful conduct.
Institutional, historical or systemic	These are witting or unwitting patterns of bias that are deeply embedded in institutions and systems of society, including the criminal justice system. Examples include racial and socioeconomic biases or disparity, which can lead to discrimination. It could also include datasets that reflect imperfect or skewed enforcement or investigation practices.

In digital criminology, when biased automated decision-making systems lead to discriminatory outcomes, the types of biases mentioned above become interlinked. However, it is also important to dissect the different biases separately to understand the wider implications, and this can be done through an interdisciplinary approach to bias. It is only by taking an interdisciplinary approach that ongoing attempts to understand, oversee, and regulate data-driven approaches within criminal justice can hope to achieve success.

Suggested reading

Babuta, A., & Oswald, M. (2019). *Data Analytics and Algorithmic Bias in Policing.* RUSI Briefing Paper. Available at https://rusi.org/explore-our-research/publications/briefing-papers/data-analytics-and-algorithmic-bias-policing.

Borgesius, F. (2018). *Discrimination, Artificial Intelligence and Algorithmic Decision-Making.* The Council of Europe. Available at https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.

Butcher, B., Robinson, C., Zilka, M., Riccardo, F., Ashurst, C., & Weller, A. (2022). Racial disparities in the enforcement of marijuana violations in the US. *AI Ethics and Society (AIES) '22 Conference, Oxford, United Kingdom, 1–3 August.* Available at https://doi.org/10.1145/3514094.3534184.

Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *N.Y.U. L. REV. ONLINE*, 94, 192–233.

References

Allen, R., & Masters, D. (2021). *Technology Managing People – The Legal Implications.* Trade Union Congress. Available at https://www.tuc.org.uk/research-analysis/reports/technology-managing-people-legal-implications (Accessed 3 July 2023).

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias, *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Accessed 3 July 2023).

Babuta, A., & Oswald, M. (2019). *Data Analytics and Algorithmic Bias in Policing.* RUSI Briefing Paper. Available at https://rusi.org/explore-our-research/publications/briefing-papers/data-analytics-and-algorithmic-bias-policing (Accessed 5 July 2023).

Baroness Casey of Blackstock. (2023) *Baroness Casey Review: An independent review into the standards of behaviour and internal culture of the Metropolitan Police Service.* Available at: https://www.met.police.uk/police-forces/metropolitan-police/areas/about-us/about-the-met/bcr/baroness-casey-review/ (Accessed 3 July 2023), p.17.

Birhane, A. (2022). The limits of fairness. *AI Ethics and Society (AIES) '22 Conference*, Oxford United Kingdom, 1–3 August. Available at: https://doi.org/10.1145/3514094.3539568 (Accessed 5 July 2023).

Bhuiyan, J. (2023). First man wrongfully arrested because of facial recognition testifies as California weighs new bills. *The Guardian*, 27 April. Available at: https://www.theguardian.com/us-news/2023/apr/27/california-police-facial-recognition-software (Accessed 3 July 2023).

Borgesius, F. (2018) *Discrimination, Artificial Intelligence and Algorithmic Decision-Making.* The Council of Europe. Available at: https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 (Accessed 5 July 2023).

Butcher, B., Robinson, C., Zilka, M., Riccardo, F., Ashurst, C., & Weller, A. (2022). Racial disparities in the enforcement of marijuana violations in the US. *AI Ethics and Society (AIES) '22 Conference*, Oxford, United Kingdom, 1–3 August (pp. 130–143). Available at: https://doi.org/10.1145/3514094.3534184 (Accessed 5 July 2023).

Centre for Data Ethics and Innovation (CEDI). (2023) *Enabling responsible access to demographic data to make AI systems fairer*, 14 June. Available at: https://www.gov.uk/government/publications/enabling-responsible-access-to-demographic-data-to-make-ai-systems-fairer (Accessed: 5 July 2023).

Centre for Evidence-Based Medicine (CEBM) (2018) *Bias – Etymology and Usage.* Available at: https://catalogofbias.org/2018/04/10/a-word-about-evidence-4-bias-etymology-and-usag/ (Accessed: 5 July 2023).

Citron, D. (2008). Technological due process. *Washington University Law Review*, 85(6), 1271–1272.

Cobbe, J. (2019) Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies*, 39, 653–654.

Davidson v Scottish Ministers [2004] UKHL 34.

Davies, B., & Douglas, T. (2022). Learning to discriminate: The perfect proxy problem in artificially intelligent criminal sentencing. In J. Ryberg and J. V. Roberts (eds.), *Sentencing and Artificial Intelligence* (pp. 97–121). Oxford: Oxford University Press.

Fogliato, F., Xiang, A., Lipton, Z., Nagin, D., & Chouldechova, A. (2021). On the validity of arrest as a proxy for offense: Race and the likelihood of arrest for violent crimes. *AI Ethics and Society (AIES) '21 Conference*, Virtual Event USA, 19–21 May.

Gerards, J., & Borgesius, F. (2022). Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence. *Colorado Technology Law Journal, 20(1), 1–55.*

Goel, S., Shroff, R., Skeem, J., & Slobogin, C. (2019). The accuracy, equity, and jurisprudence of criminal risk assessment. In R. Vogl (ed.), Research Handbook on Big Data Law (pp. 9–28). Cheltenham: Edward Elgar Publishing.

Johnson, K. (2022). How wrongful arrests based on AI derailed 3 men's lives. *Wired*, 7 March. Available at: https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/ (Accessed 5 July 2023).

Kaufmann, M. (2023). *Making Information Matter.* Bristol: University of Bristol Press.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores *Innovation in Theoretical Computer Science (ITCS) 2017 Conference, Berkeley, California, USA, 9–11 January.* Available at: https://drops.dagstuhl.de/opus/volltexte/2017/8156/pdf/LIPIcs-ITCS-2017-43.pdf (accessed 5 July 2023).

Magnet, S. A. (2011). *When Biometrics Fail: Gender, Race, and the Technology of Identity.* Durham, NC: Duke University Press.

Mansfield, T. (2023). *Facial Recognition Technology in Law Enforcement Equitability Study: Final Report.* National Physical Laboratory. Available at: https://science.police.uk/site/assets/files/3396/frt-equit ability-study_mar2023.pdf (Accessed 5 July 2023).

Monteith, K., Quinn, E., Joseph-Salisbury, R., Dennis, A., Kane, E., Addo, F., & Mcgourlay, C. (2022). *Racial Bias and the Bench: A Response to the Judicial Diversity and Inclusion Strategy.* University of Manchester. Available at: https://documents.manchester.ac.uk/display.aspx?DocID=64125 (accessed: 16 May 2023), pp. 8, 14, 17.

Oswald, M., Grace, J., Urwin, S., and Barnes, G. (2018) Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'Experimental' proportionality. *Information & Communications Technology Law*, 27(2), 223–250.

R v Gough [1993] 2 WLR 883.

Richardson, R., Schultz, J., and Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *N.Y.U. L. REV. ONLINE*, 94, 192–233.

Rieke, A., & Bogen, M. (2018). *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias.* Upturn. Available at: https://www.upturn.org/work/help-wanted/ (Accessed 3 July 2023), p. 9.

Sanchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives. *FAT '20 Conference, Barcelona, Spain, 27–30 January,* 458–468.

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Bias preservation in machine learning: The legality of fairness metrics under EU non-discrimination law. *West Virginia Law Review*, 3(123), 735–790.

Aleš Završnik

12 Big data

Abstract: The chapter presents the changing meaning of Big data over time and its relation to digitalization in contemporary 'datafied' society, in which Big data has become viewed as the 'new oil.' The chapter then delves into the question of relevance for criminology. It presents different views and framings of its benefits and risks in the crime control domain and for the production of criminological knowledge. It presents typical uses of Big data in crime control practice and some of the risks to fundamental liberties theorof. Connections with the notion of Artificial Intelligence, narratives of risk, prediction, and pre-emption are also presented.

Keywords: Big data, datafication, human rights, production of knowledge, prediction, pre-emption

Big data gained social science relevance due to increased data processing practices by public institutions and private companies that involve large amounts of data about citizens and (potential) customers. 'Big' data, as opposed to 'small' data, was perceived as a game changer for analyzing and understanding the social realm and for the design of policies, such as health policy, social policy, macroeconomic policy, and also crime policy. Ultimately, the trend was dubbed as a Big data "revolution" (Lavorgna and Ugwudike, 2021). The term 'Big data' is used to describe large new data sources, as well as the associated processes, i. e., collection, analysis, insight, and Big data 'mining'—a process of uncovering patterns in large datasets. Critical authors define Big data not only as the capacity to search, aggregate, and cross-reference large datasets but also as a "specific socio-technical phenomenon" (boyd and Crawford, 2012: 663).

Big data plays an increasingly large role in fields relevant to criminologists. Data gathered from several sources is claimed to offer new perspectives and insights into reasons, factors, and circumstances of past and potential future crimes. The expectation towards Big data is high: not only would big datasets offer new understandings of crime but implicitly also new approaches on how to act upon these, whether it is to prevent crime (e. g., with predictive policing software) or investigate already committed crime (e. g., with a prediction of a risk score of parolee). Its biggest promise seems to be to predict future crime (e. g., place, actors—perpetrators and victims) in order to pre-emptively 'strike' and prevent crime from happening, i. e., to colonize the future (see Prediction by Ķīlis, Gundhus and Galis). The fact that Big data has its origins in the business world has ramifications for the crime control domain, especially in pur-

Acknowledgments: Aleš Završnik acknowledges receiving financial support from the Slovenian Research and Innovation Agency (research core funding no. P5 – 0221 and the research project "Proteverb – Legal, ethical and technological aspects of processing textual and speech data for scientific, research and development purposes", No. V5 – 2265).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-013

suing values of effectiveness in a manner of 'doing more with less.' It is linked to the neoliberal turn when shrinking police budgets meant that the police must ensure the same level of protection with limited resources (Beck and McCue, 2009).

The concept of Big data is relatively new, but the origins can be traced back to the 1960s and '70s with the rise of data centers and relational databases. Around 2005, with the advent of social media platforms, it was clear that users were generating data that could be 'monetized' and made 'actionable' (Zuboff, 2015). In the 2010s, Big data had become a topic of discussion in various domains, and authors were showing enthusiasm that parallels some of the most significant movements in the history of computing, such as the development of personal computing in the 1970s, the World Wide Web in the 1990s, and social media in the 2000s.

The 'big' in Big data has acquired different meanings over time, and 'data' in Big data generated its own field of research in the social sciences, including software studies and critical data studies. What seemed 'big' a decade ago became 'small' with the exponential growth of technological capacities. A telling example comes from increasingly powerful Large Language Models (LLM), which are built on supposedly the 'whole internet.' For instance, the ChatGPT model required a data center with 10,000 GPUs, i. e., Graphics Processing Units, which are electronic circuits that can perform high-calibre mathematical calculations. ChatGPT training lasted 9 months and cost $100 million in electricity. In 2023, a new data centre with 50,000 GPUs was built, which meant that the cost of the processors alone approached $2 billion (Zgonik, 2023). The volume of 'big' has skyrocketed due to more and more data generated by digital appliances—from phones, wristwatches, to fridges, heating systems, and other objects of mundane life connected to the Internet of Things (IoT) or the 'Internet of everything" (see Internet of Things by Milivojevic). Some have hence also claimed the opposite—that 'small' quality data rather than 'big' 'anything' data is something to aspire to in order to acquire actionable insights to be acted upon.

Today, Big data should also be conceptualized together with other technologies, forming the *Big data pipeline* or *Artificial Intelligence (AI) supply chain.* 'Ingredients' (Big data) do not lead to meaningful outputs without a proper 'recipe' (AI). Since the meaning of large amounts of data must be extracted in order to furnish 'actionable' insights, AI is an essential part of the Big data pipeline. The two—Big data and AI—are related as part of the same logic and values.

Criminological relevance

The first explicit recognition of the relevance of Big data for criminology can be traced to the 2010s when Berk (2012) analyzed changes in computational criminology brought about by machine learning. Forecasting in criminal justice, together with predictive policing, were two specific applications of Big data for crime control. Procedures from computer science and applied mathematics have been used before the 'advent' of Big data to animate theories about crime and law enforcement, but a culture of causal

modeling thoroughly dominated such methods. Machine learning, instead, comes from a different culture characterized by an 'algorithmic perspective' (Berk, 2013).

Big data entered into criminology through several paths: through the increasing use of new types of data (social media or user-generated data), through using computer modeling/algorithms as a predictive tool to guide policing strategies and other criminal justice decisions (Chan and Bennett Moses, 2016) and also through the theorization of the increased focus on risk (Zedner, 2007). The latter is part of wider skepticism of Big data in criminology and shows how criminologists have scrutinized the meaning, circulation, and power relations associated with Big data. Authors reflected on whether the claim that the 'data deluge' would make scientific methods obsolete has merits for criminology (Chan and Bennett Moses, 2016) and analyzed predictive policing programmes (Egbert and Leese, 2021). Critical criminology challenged the very novelty of Big data and claimed that the digital turn has failed to fulfill the old dreams of more just and equal societies (Završnik, 2018). What may be new with the Big data 'revolution' is the exponential acceleration of neoliberalism, for example reflected in reduced state power and a reinforced private sector, an increase in social and wealth inequalities facilitated by the powerful elite to gain insight into different populations more than ever before. The pressure to institute a process of 'datafication'—turning everything into data or numbers—in order to 'monetize' data and create 'actionable' insights is at the core of Big data logic (see Datafication by Chan). Big data, hence, serves specific political ends. Similar to Desrosières' (2002) analysis of statistics, for which he lucidly noted that it offered new justifications for modern state interventions back in the 19th century, today, Big data offers new justifications for policy interventions. Big data is, hence, in this sense, not an 'objective' knowledge but has always been a political endeavor.

Other social sciences were also relevant for criminological attempts to grasp the Big data 'revolution.' As a multidisciplinary field of research, criminology has drawn knowledge from critical data studies and critical security studies, critical media studies, critical legal studies, and the work of many authors studying at the intersection of political science, sociology, Science and Technology Studies (STS), ethics, and surveillance studies.

Uses of Big data in crime control

Examples of uses of Big data in the crime control domain include:

1) predictive policing software, traditionally using Big data to predict where and when crimes are likely to occur. The focus is typically on locations ('hot spots') or individuals ('heat lists,' 'persons of interest') (e. g., Kaufmann et al., 2019), which is the case in the software PRECOBS (*Pre Crime Observation System*) which creates graphically supported insights into areas where increased risk for follow-up incidence is estimated within the next 72 hours. Another example is *HunchLab*, now Azavea, a predictive policing program that inputs crime data, cen-

sus data, population density and adds variables like the location of schools, churches bars, clubs, and transportation centers (Ferguson, 2017);
2) social network analysis to map the relationships between criminals and their associates, to identify potential suspects, and analyze criminal networks;
3) DNA databases and other biometric databases for profiling (Kaufmann, 2022);
4) analytic tools for criminal investigations, e. g., to extract knowledge on criminal networks from multiple data sources in real-time, gunshot detection, or video analysis of child abuse images;
5) analytic tools to assist bail courts, e. g., with generating risk scores in bail procedures and assisting parole bodies (Bennett Moses and Chan, 2014).
6) sentencing tools to predict an offender's likelihood of future recidivism, e. g., COMPAS (*the Correctional Offender Management Profiling for Alternative Sanctions*, used in the USA) or *the Risk of Reconviction* (in the UK) and the LSI (*Level of Service Inventory*, used internationally) (Ryberg and Roberts, 2022);
7) tools for legal reasoning and legal research (Legal-tech).

Challenges of Big data in crime control

Since data are the focal point of the Big data paradigm, *privacy and surveillance concerns* form an overarching human rights issue (Kerr and Earle, 2013). Increasing collection and processing capabilities of Big data analytics change *surveillance* of daily life and can have a chilling effect on free speech and other civil liberties. The relation of Big data to surveillance has been theorized as a distinct type of Big data surveillance (Andrejevic and Gates, 2014) and dataveillance. Reflecting on Snowden's revelations of 'dragnet; investigation practices by intelligence agencies, Lyon (2014) observed a transition in surveillance studies from information technology and networks to Big data, which intensified and expanded certain surveillance trends. The future orientation and the quest for pattern discovery of Big data surveillance raised concerns related to privacy, social sorting, and pre-emption (Lyon, 2014).

Data on identifiable individuals collected in legally and/or ethically problematic fashion, such as social media scraping, remain part of Big datasets with unclearly defined data subjects' rights. For instance, it is unclear whether and how data subjects can exercise the rights guaranteed in the personal data protection regime vis-à-vis facial recognition technologies (see Facial Recognition by Fussey).

The *quality of data*, relating to (in)accuracy, completeness of data, representativeness of social groups, etc., varies in the criminal justice domain. Which data is taken in, and which data is left out of the calculus? Data is not a natural resource but a cultural one, it is always 'baked' with (underpinning) human values, interests, and cultural expectations (Gitelman, 2013). Criminal justice data *per definition* does not include unreported crimes (dark figure of crime), which can make it difficult to prevent the so-called 'garbage in—garbage out' effect: poor data leads to poor outcomes. Big datasets, then, tend to indicate social practices (reporting crimes) rather than social reality.

One aspect of data quality relates to *biased data*, which may lead to *discrimination* (see Bias by Oswald and Paul). Research on sentencing prediction instruments has confirmed how criminal history is, in fact, a proxy for race (Harcourt, 2015), meaning that data for such instruments is biased by racialized histories. Police data often reflects biased functioning of police operations, and Big data analytics would perpetuate such biases. If digitized societies are divided along gender, race, wealth, and other lines, and structural inequalities (Ávila et al., 2019: 97), practices based on Big data will *per definition,* reflect such biases (called 'closed-loop'). Examples abound and include the over-policing of minority neighborhoods and the under-policing of white-collar crime as predictive policing tools focus on street and property crimes. As O'Neil (2016) vividly expressed, models are opinions embedded in math. The interpretation of results of Big data analytics is not straightforward as well. It is inherently affected by human knowledge of the analyzed domain and data. Data scientists must work alongside domain-specific scientists in order to ascribe meaning to the calculated results.

The *cost* is another concern, as Big data tools are increasingly sophisticated and expensive to train and maintain, e.g., Microsoft and OpenAI used $100 million worth of energy for training ChatGPT, and they keep spending $700k per day for running it (Zgonik, 2023).

The *'Black box' effect* and a lack of *transparency* are general problems of Big data analytics (Pasquale, 2015). Criminologists have also raised this concern over the latest fifth generation of machine learning-based (ML) risk assessments. Ávila et al. (2021) claim that while the latest ML-based risk assessments are focused on the elimination of biases and self-adjust to new data over time, they also deepen the black box problem. They claim that opacity, proprietary nature, and fluid characteristics of predictive models undermine legal protections.

Blurring regulatory boundaries is another critical aspect of Big data use in crime control. The new mathematical language serves security purposes well (Amoore, 2014). Here, new concepts are being invented in order to understand crime (knowledge production) and act upon it (crime control policy). Such concepts include 'meaning extraction,' 'sentiment analysis,' and 'opinion mining.' However, these concepts are blurring the boundaries in the crime control domain: instead of the relatively well-defined concepts of criminal law, such as suspect, reasonable doubt, etc., which serve as regulators of and thresholds for the intervention of law enforcement agencies, new concepts no longer sufficiently confine agencies nor prevent abuses of power (Završnik, 2021).

Risk, prediction, and pre-emption

A central use of Big data analytics in criminology is to garner predictions and identify risks with the final goal of managing and pre-empting the risks. Anticipation re-focuses crime control actors (Zedner, 2007). They reorient their practices and "focus on the fu-

ture more than on the present and the past. In the context of neo-liberal governance, this anticipation is likely to place more weight on surveillance for managing consequences rather than research on understanding causes of social problems such as crime and disorder." (Lyon, 2014: 6–8). Big data, so the critique, tends to accompany this shift from causation (e.g., aiming to uncover factors leading to crime) to correlation (e.g., aiming to uncover factors of crime that need to be 'managed' here today).

Pre-emptive approaches in the crime and security domain have been growing steadily since the 1990s and have been extensively augmented after 9/11. Such approaches are a bureaucratic incentive to over-collect data (Lyon, 2014). Big data is at the core of the transition toward pre-emptive approaches in tackling crime. It enabled prediction and triggered a new philosophy of *pre-emption.* Predictive analytics that transcends human perception have been one of the most attractive aspects regarding the application of Big data in crime control, for example 'connecting the dots' in terabytes of data in money laundry schemes would be impossible for a human eye, while Big data analytics can help follow the money by finding hidden correlations. However, Kerr and Earle (2013) warn that Big data's promise of increased efficiency, reliability, and utility might be seen as the justification for a fundamental jurisprudential shift from an *ex-post facto* system of penalties and punishments to *ex-ante* preventative measures. The new form of 'pre-emptive prediction,' as Kerr and Earle (2013) define it, is intentionally used to diminish a person's range of future options. Predictions are used to assess the likely consequences of allowing or disallowing a person to act in a certain way (Kerr and Earle, 2013). Predictions, here, are not concerned with an individual's actions but with whether an individual or group should be permitted to act in a certain way. Big data is thus used "not only to understand a past sequence of events, but also to predict and intervene before behaviours, events, and processes are set in train (sic) [i.e. motion]" (Lyon, 2014: 4). Pre-emption means acting to prevent an anticipated event from happening. Taken to its extreme, the philosophy of pre-emption is not merely pro-active—it is aggressive. Kerr and Earle (2013) exemplify that no-fly lists employing predictive algorithms curtail liberties. Before their development, high-risk individuals were generally at liberty to travel unless the government had sufficient reason to believe that such individuals were in the process of committing an offense. But now, a no-fly list obliterates the need for such evidence. Prediction replaces the need for proof. Big data underwrites anticipatory and pre-emptive approaches that move crime policy towards actuarialism and consequentialist concerns with managing crime rather than seeking its causes in an attempt to eliminate them.

Conclusion

The discussions on Big data remain highly pertinent to contemporary criminology, albeit sometimes being framed as a discussion on risk, prediction, prevention and pre-emption, automation and 'algorithmization' of crime and crime control. AI tools should be understood as part of the 'Big data pipeline.' Analysis of automated decision-making

in criminal justice generating new forms of 'automated justice' (Marks et al., 2015), 'algorithmic justice' (Završnik, 2021), 'simulated justice' (O'Malley, 2010) should be conceptualized together with findings of Big data studies. Big data is then part of the wider trend of 'algorithmic governance' and 'algorithmic governmentality' (Hannah-Moffat, 2019). It should also be read together with critiques that were transitioning from 'the rule of law' to 'the rule of algorithms' ('algocracy').

Big data can *enhance* criminology's scientific method in understanding patterns of crime and analyzing and verifying theories of crime. Big data analytics may also improve the effectiveness, legitimacy of criminal justice actors and increase the investigative powers of law enforcement agencies. However, a *critique* of Big data focusing on Big data's contribution to reducing democratic freedoms, reconfiguring privacy and redefining the role of information in contemporary societies needs to accompany the implementation of Big data tools. A discussion on the limits and thresholds in the use of Big data analytics in crime and social control is needed, one of which is, for example, bulk biometric surveillance. The change in orientation of traditional criminal justice based on an after-the-fact system of punishments to one based on future-oriented preventative measures—anticipation and pre-emption induced by Big data must be examined and regulated. The need for *cross-disciplinary dialogue* between developers, data scientists, analysts, criminologists, and others about the legal, socio-political, and discriminatory effects of Big data analytics cannot be understated.

Suggested reading

Amoore, L. (2014). Security and the incalculable. *Security Dialogue*, 45(5), 423 – 439.

Chan, J., & Bennett Moses, L. (2016). Is Big Data challenging criminology? *Theoretical Criminology*, 20(1), 21 – 39.

Hannah-Moffat, K. (2019). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, 23(4), 453 – 470.

Završnik, A. (ed.) (2018). *Big Data, Crime and Social Control*. Abingdon and New York: Routledge

References

Amoore, L. (2014). Security and the incalculable. *Security Dialogue*, 45(5), 423 – 439.

Andrejevic, M., & Gates, K. (2014). Big Data surveillance: Introduction. *Surveillance & Society*, 12(2), 185 – 196.

Ávila, F., Hannah-Moffat, K., & Maurutto, P. (2021). The seductiveness of fairness: Is machine learning the answer? – Algorithmic fairness in criminal justice systems. In M. Schuilenburg & R. Peeters (eds.), *The Algorithmic Society. Technology, Power, and Knowledge* (pp. 87 – 103). Abingdon & New York: Routledge.

Beck, C., & McCue, C. (2009). Predictive Policing: What can we learn from Wal-Mart and Amazon about fighting crime in a recession? *The Police Chief Magazine*, 76(11), 18 – 24.

Bennett Moses, L., & Chan, J. (2014). Using big data for legal and law enforcement decisions: Testing the new tools. *University of New South Wales Law Journal*, 37(2), 643 – 678.

Berk, R. (2012). *Criminal Justice Forecasts of Risk: A Machine Learning Approach.* New York: Springer.
Berk, R. (2013) Algorithmic criminology. *Security Informatics*, 2(1).
boyd, d., & Crawford, K. (2012). Critical questions for big data' *Information, Communication & Society*, 15(5), 662 – 679.
Chan, J., & Bennett Moses, L. (2016). Is Big Data challenging criminology? *Theoretical Criminology*, 20(1), 21 – 39.
Desrosières, A. (2002). *The Politics of Large Numbers: A History of Statistical Reasoning.* Cambridge, MA: Harvard University Press.
Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* Abingdon & New York: Routledge.
Ferguson, A. G. (2017). *The Rise of Big Data Policing. Surveillance, Race, and the Future of Law Enforcement.* New York: New York University Press.
Gitelman, L. (ed.) (2013). *'Raw Data' is an Oxymoron.* Cambridge, MA: MIT Press.
Hannah-Moffat, K. (2019). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, 23(4), 453 – 470.
Harcourt, B. E. (2015). Risk as a proxy for race. *Federal Sentencing Reporter*, 27(4), 237 – 243.
Kaufmann, M. (2022). DNA as in-formation. *WIREs Forensic Science.* Online First.
Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.
Kerr, I., & Earle, J. (2013). Prediction, preemption, presumption: How big data threatens big picture privacy. *Stanford Law Review Online*, 66(65), 65 – 72.
Lavorgna, A., & Ugwudike, P. (2021). The datafication revolution in criminal justice. *Big Data & Society*, 8(2).
Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 1 – 13.
Marks, A., Bowling, B., & Keenan, C. (2015). Automatic justice? Technology, crime and social control. In R. Brownsword, E. Scotford, & K. Yeung (eds.), *The Oxford Handbook of Law, Regulation and Technology* (pp. 705 – 730). Oxford: OUP.
O'Malley, P. (2010). Simulated justice: Risk, money and telemetric policing. *British Journal of Criminology*, 50(5), 795 – 807.
O'Neil, C. (2016). *Weapons of Math Destruction.* New York: Crown.
Pasquale, F. (2015). *The Black Box Society.* Cambridge, MA: Harvard University Press.
Ryberg, J., & Roberts, J.V. (eds.) (2022). *Sentencing and Artificial Intelligence.* New York: Oxford University Press.
Završnik, A. (ed.) (2018). *Big Data, Crime and Social Control.* Abingdon and New York: Routledge.
Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 18(5), 623 – 642.
Zedner, L. (2007). Pre-crime and post-criminology? *Theoretical Criminology*, 11(2), 261 – 281.
Zgonik, S. (2023, 8 July). *Dr. Jure Leskovec za N1.* N1 News.
Zuboff, S. (2015). Big Other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75 – 89.

Ravida Din and Shoshana Magnet

13 Biometric failure

Abstract: This definition explores how biometric technologies continue to codify a cultural context rife with systemic inequalities.

Keywords: surveillance, security studies, feminist theory, algorithmic bias

Defining biometrics

Biometric technologies render bodies in binary code. Reducing complex and messy material bodies to a series of zeroes and ones—biometric technologies are aimed at making bodies more easily identifiable, speeding the passage of people and commerce. In the words of scholar Katja Franko Aas, biometric and other digital technologies aimed at reading the body "mean that [b]odies become 'coded' and function as 'passwords'" (2006). Biometrics take an electronic reading of the body, an identification process that communication scholars note is at the intersection of multiple communications and information processes, including photography, videography, computer networking, pattern recognition, and digitalization (Gates, 2011). Biometric technologies include, but are not limited to, iris scanning, retinal scanning, digital fingerprinting technologies, and facial recognition technologies. Biometric technologies were originally developed for the prison industrial complex and tested coercively on a population of prisoners that could not refuse to use them. Unsurprisingly, given their beginnings, biometric technologies do not work as claimed for them in ways intimately related to existing inequalities. For example, biometric facial recognition technologies have had and continue to have difficulty reading the bodies of racialized, queer, and disabled people as well as bodies in professions that require a lot of handwashing such as nurses and teachers (Magnet, 2011). In another example, since their onset, retinal scanners have had difficulty identifying the irises of people with dark brown eyes (Magnet, 2011). First used in the prison industrial complex, biometric technologies are now staples of identification in a number of institutions, including state welfare systems, immigration and refugee centers as well as being frequently used in consumer technologies.

How do we define biometric technologies? Definitions are powerful. In Paula Treicher's study of HIV/AIDS, she tells us that there were high stakes involved in naming and definitions within the epidemic, not only for "patent rights to the lucrative test kits for the AIDS virus … but the future and honor of immunology" (1999: 31). So too with biometric technologies. How biometrics are defined concerns not only corporations, policymakers, and consumers, but those who have been forcibly subject to biometric scans in ways intimately connected to what bell hooks called "white suprema-

Acknowledgments: This project is supported by a SSHRC Insight Development Grant # 435 – 2021 – 0035.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-014

cist capitalist heteropatriarchy" (hooks, 1999). Are biometrics a consumer technology that clients may opt-in to in order to facilitate transactions? Are they a requirement of being able to hold a travel document or to get a job? Biometric companies hoping for as many clients as possible have fought to define biometric technologies as an all-purpose, privacy-enhancing solution to the problem of security in the age of insecurity. And yet, so too have biometric technologies been used by privacy and human rights organizations as the case study *par excellence* that reveals the dangers of how technologies can destroy individual and group privacy rights (see Privacy and Data Protection by Bygrave). While "meanings may be multiple, and, even if contradictory, allowed to co-exist in a single space or 'a single head'" (Treichler, 1999), definitions are "less democratic," representing the "outcome of a struggle." How might we complexify the definition of biometric technologies? We argue that a place to start would begin by being attentive to the ways that biometric systems break down or fail.

On failure and biometrics

Failure of biometric technologies to work in the ways claimed for them entrench existing forms of racism, sexism, and ableism as well as homo- and trans- phobias (Magnet, 2011; Browne, 2015). For example, they reliably fail to work distinguishing the faces of people of color (as US and Canadian biometric scanners screening people at the border had difficulty locating the faces of people of color, or had trouble locating the irises of people with dark brown eyes), people with disabilities (as the biometric scanners were too tall for wheelchair users or for little people, but too short for unusually tall people), and queer people (as biometric facial recognition technology is used to split people up into gender categories that rely on assumptions that women have long hair and that only men have short hair and only men wear ties) (Magnet, 2011; Dubrofsky and Magnet, 2015; Hoffer and Savini, 2017; Benjamin, 2019; Weitzberg, 2020). That is, biometric technologies are "haunted" (Gordon, 1997) by histories of white supremacy that assume racialized bodies to be impossibly alike or "haunted" by sexist claims that bodies can speak the "truth" of their gender identities. In other words, biometric technologies are used to perpetuate the myth that the body is no more than "a password" that speaks the truth of its identity (Aas, 2007). (This myth persists despite 100 years of failed attempts to get a lie detector test to work in a reliable fashion; Littlefield, 2009; Cole, 2002). Biometric companies attempt to imagine bodies as stable entities that can give us the definitive proof of identity in ways that permit state actors—including the corporations and governmental departments involved in the prison industrial complex—to surveil and regulate othered communities and populations. And yet, biometric mismatches due to mechanical failures and the technology's inability to work objectively dispute such stability.

Biometric technologies and their attendant failures reveal how we continue to have an impoverished language for thinking more broadly about technological failure. In just one example of this lack, why don't we think of the ways that new technologies

might contribute to the intensification of existing inequalities as failures? For example, biometric technologies rely upon erroneous assumptions about the biological nature of race, gender, and sexuality (Magnet, 2011; Browne, 2010). In so doing they result in bodies that cannot produce a biometric image—whether fingerprint, faceprint, or retinal scan—resulting in individuals who are denied their basic human right to mobility, employment, food, and housing if they cannot be identified by the biometric scanner. Although biometric scientists often speak of 'false accept' or 'false reject' biometric errors, we continue to lack language for thinking about the failures of biometric technologies to contribute to substantive equality. As authors Kaufmann and Vestad remind us, "Criminology has an intimate link to visualization" (2023). So too are the failures of the prison industrial complex intimately bound up with the failures of biometric technologies to visualize the very bodies they claim to be able to codify, as in the case of bodies that simply cannot produce a biometric fingerprint or a reliable retinal scan (Browne, 2010). Moreover, these failures are not simply technical failures, in that biometric technologies are used to visualize crime and criminality in ways that intensify the criminalization and surveillance of racialization from the border to the prison—as well as used to strengthen existing systems of crimmigration—or the criminalization of immigration (van der Woude et al., 2017). As Kaufmann and Vestad argue: "Visual profiles of bodies never emerge out of nowhere. They are a socio-technical practice. They are embedded in the history of explaining crime and the scientization of police and legal work" (Kaufmann and Vestad, 2023).

Biometric technologies are part of the larger problem of technologies that intensify algorithmic bias (Bias by Oswald and Paul), from robotics to search engines to new identification technologies. In the field of algorithmic bias, there is a tendency to think of algorithms as independent entities that are unencumbered by human reflexivity. Algorithms are imagined to hold the promise of objectivity (Daston and Galison, 2007), an objectivity that is perniciously sold to us by large corporations eager to have markets for their technological products. The persistence of "technochauvinism," defined by Meredith Broussard as "a kind of bias that considers computational solutions to be superior to all other solutions" (2023: 2), reminds us how biometric technologies can be sold to consumers as replacing the subjective human eye with the objective gaze of the scanner—failing to think about how the scanner itself codifies existing systemic discrimination.

Artificial Intelligence (AI) encompasses a broad set of algorithms alongside its scaffolding of 'deep learning' and 'machine learning' systems (see Artificial Intelligence by Van Brakel). Central to the operation of these digital technology systems is the availability of massive datasets. According to John Cheney-Lippold, "… one of the primary terms of an algorithm is that everything is represented as data" (2017: 11). It is reasonable, therefore, to assume that algorithms are only as good as the data they are fed, remembering that data is neither made nor collected in a lab sterile to culture. Scholars Safiya Noble (2018), Ruha Benjamin (2019), Wendy Hui Kyong Chun (2021), and Katherine McKittrick (2021) all show that technology is made by human hands—one that are steeped in both a history and a present of white supremacy, of misogyny, of homo-

phobia and of other systemic forms of discrimination. These authors—who are part of larger conversations about algorithmic bias—theorize how historical data, steeped in a broader social context of systemic inequality, leads to discriminatory outcomes that preserve the status quo. For example, describing this reality as the "New Jim Code," Ruha Benjamin argues that this "imagined objectivity reflects and reproduces social hierarchies, whether wittingly or not" (2019: 2) whereas Safiya Noble (2018) describes the ways that search engines encode an existing context of white supremacy as "algorithmic oppression" (p. 17). Furthermore, Joy Buolamwini's groundbreaking work on facial recognition technology too shatters the myth of race as absent from technology when she discovered that the software that she was using failed to recognize her until she donned a white mask. In other words, the machine rendered blackness as illegible, reifying whiteness as its default (Buolamwini and Gebru, 2018). Google's photo-app system, lauded as a pinnacle of human identification, consistently labeled Black people as gorillas. An insidious dehumanization produced through computational 'detachment' is, as Katherine McKittrick explains, "tied to a biocentric system of knowledge" (2021: 111).

Algorithmic bias and complex sorrows

By now, it is well established that algorithmic 'thinking' which produces 'mistakes-to-be-corrected' are not without consequence. The claim that these are simply 'glitches' in the system that require a technical reconfiguration runs contrary to the fact that algorithmic decision-making remains invisible (commonly referred to as the 'black-box' syndrome) to us. Irene Fubara-Manuel (2019) draws on Simone Browne's concept of "critical biometric consciousness" to underscore a praxis of accountability. She writes, "this accountability demands that the architects of these technologies consider the inscriptive practices of biometrics in placing identity on the body. More so, it requires accountability for the ways in which these technologies further subjugation based on race, gender, and disability" (p. 70).

Important to the definition of biometrics is an extension of Marianne Paget's work on mistakes as complex sorrows. Marianne Paget (1993), a sociologist of medicine, argues that "clinical medicine is an error-laden activity" and that we lack words to think about medical mistakes in a complex way. Instead, language taken from legal and insurance discourses about fault and blame dominate our current ways of speaking about medical error; thus, we use the terms *medical malpractice*, *misdiagnosis*, and *negligence*, words that do not get at the affective or emotional nature of medical 'mistakes,' a word seldom used and considered imprecise for the purpose of insurance and legal claims. Biometrics uses the scientific language of 'false acceptance rates' and 'failure to enroll rates' to describe biometric failures. These failures are often understood as exceptional, the result of 'a few bad apples,' rather than as endemic to the science of identification itself. And yet, when biometric scanners are not recognizing the irises of people with dark brown eyes, when retinal scanning technologies are not made acces-

sible to wheelchairs users, and when facial recognition technology disproportionately misrecognizes racialized faces, we know that these errors are endemic to the technologies themselves (Fubara-Manuel, 2019; Magnet, 2011). Given the high stakes of biometric identification, our desire for an error-free science of recognition is great. Yet the desire for error-free biometric technologies does not magic them into existence. Paget refers to medical errors as "complex sorrows" to help us think about both the errors' scientific and emotional implications. Given the devastating consequences of biometric errors for human lives, we need a language that is not restricted to technical terms (see Error by Aradau). Biometric failures as complex sorrows are a beginning.

As I [Shoshana] wrote about in *When Biometrics Fail: Race, Gender and the Technology of Identity* in 2011:

Biometrics became useful as a form of what the security technologist Bruce Schneier (2006) has termed 'security theatre,' or what the border theorist Peter Andreas (2003) calls a 'politically successful policy failure.' Regardless of how often the technologies broke down, worked differentially depending on race and gender, or cost rather than saved the state money, biometrics were used to suggest that *something* was happening. Thus, even when biometric technologies failed, they also succeeded. Whether they were used to assert that the state was getting tough on crime, welfare, or terrorism, biometric technologies were offered up as proof that public–private partnerships were working for the public good, whatever the particular good of the moment might be. Yet despite persistent mechanical failures, biometric technologies still accomplish a great deal for state and commercial actors whose interests are tied to contemporary cultures of security and fear. In this sense biometric technologies succeed even when they fail. On the other hand, even when they function technically, biometrics do real damage to vulnerable people and groups, to the fabric of democracy, and to the possibility of a better understanding of the bodies and identities these technologies are supposedly intended to protect.

And in this sense, and this truth has endured, they fail even when they succeed in reliably identifying bodies with precision. Having detailed the ways that biometric technologies fail in ways intimately connected to systemic oppression—whether they fail to identify the faces of racialized people, whether they are inaccessible for people with disabilities to use at all, or whether they simply fail to work as they are described—we have to remember that they are also an important part of the performative nature of the security industrial complex. That is, biometric technologies continue to make significant profits for the corporations that sell them while suggesting that something by the way of security is indeed happening. Is this the future we want? One of increasingly entrenched technological inequality? Surely not. We need to envision different futures—ones that are not technologically reducible to a series zeroes and ones—but futures concerned with human thriving, with human complexity, and with substantive and intersectional equality. We reduce bodies to binary code at the peril of losing all of the messy complexities that make us truly human.

Suggested reading

Bean, H., & Fischer, M. (2021). Queering intelligence studies. *Intelligence and National Security*, 36(4), 584 – 598. DOI: https://doi.org/10.1080/02684527.2021.1893078

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness.* Durham, NC: Duke University Press.

Kaufmann, M., & Vestad, M. (2023). Biology and criminology: Data practices and the creation of anatomic and genomic body 'types'. *Critical Criminology*, 31, 1217 – 1232. https://link.springer.com/article/10.1007/s10612-023-09732-6

References

Aas, K. F. (2006). 'The body does not lie': Identity, risk and trust in technoculture. *Crime, Media, Culture: An International Journal*, 2(2), 143 – 158. https://doi.org/10.1177/1741659006065401

Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code.* Cambridge: Polity.

Broussard, M. (2023). *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech.* Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/14234.001.0001.

Browne, S. (2010). Digital epidermalization: Race, identity and biometrics. *Critical Sociology*, 36(1), 131 – 150. https://doi.org/10.1177/0896920509347144.

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness.* Durham, NC: Duke University Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1 – 15.

Chenney-Lippold, J. (2017). *We Are Data: Algorithms and the Making of Our Digital Selves.* New York: NYU Press.

Chun, W. H. K. (2021). *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition.* Cambridge, MA: MIT Press.

Cole, S. A. (2002). *Suspect Identities: A History of Fingerprinting and Criminal Identification.* Cambridge, MA: Harvard University Press.

Daston, L., & Galison, P. (2007). *Objectivity.* Princeton, NJ: Princeton University Press.

Dubrofsky, R., & Magnet, S. (2015). Introduction: Feminist surveillance studies: Critical interventions. In *Feminist Surveillance Studies* (pp. 1 – 18). Durham, NC: Duke University Press.

Fubara-Manuel, I. (2019). *Animating Opacity: Race, Borders, and Biometric Surveillance.* [Doctoral Thesis, University of Sussex, 2019). http://sro.sussex.ac.uk/.

Gates, K. (2011). *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance.* New York: NYU Press.

Gordon, A. (1997). *Ghostly Matters: Haunting and the Sociological Imagination.* Minneapolis: University of Minnesota Press.

Hoffer, K. J., & Savini, G. (2017). Effect of gender and race on ocular biometry. *International Ophthalmology Clinics*, 57(3), 137 – 142. https://doi.org/10.1097/IIO.0000000000000180.

hooks, b. (1999). *All About Love: New Visions.* New York: William Morrow.

Kaufmann, M., & Vestad, M. (2023). Biology and criminology: Data practices and the creation of anatomic and genomic body 'types'. *Critical Criminology*, 31, 1217 – 1232. https://link.springer.com/article/10.1007/s10612-023-09732-6

Littlefield, M. (2009). Constructing the organ of deceit: The rhetoric of fMRI and brain fingerprinting in post-9/11 America. *Science, Technology, & Human Values*, 34(3), 365 – 92. https://doi.org/10.1177/0162243908328756.

Magnet, S. A. (2011). *When Biometrics Fail: Gender, Race, and the Technology of Identity.* Durham, NC: Duke University Press.

McKittrick, K. (2021). *Dear Science and Other Stories.* Durham, NC: Duke University Press.
Nakamura, L., & Chow-White, P. (2013). *Race After the Internet.* New York: Routledge.
Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5.
Paget, M. (1993). *Complex Sorrow: Reflections on Cancer and an Abbreviated Life.* Philadelphia: Temple University Press.
Treichler, P. (1999). *How to Have Theory in an Epidemic.* Durham, NC: Duke University Press.
van der Woude, M., Barker, B., & van de Leun, J. (2017). Crimmigration in Europe. *European Journal of Criminology*, 14(1), 3 – 6. https://doi.org/10.1177/1477370816639814
Weitzberg, K. (2020). Biometrics, race making, and white exceptionalism: the controversy over universal fingerprinting in Kenya. *The Journal of African History*, 61(1), 23 – 43. https://doi.org/10.1017/S002185372000002X.

Julien Jeandesboz

14 Borders and border control

Abstract: The chapter examines how borders and border control matter for digital criminology. It first probes the relation between border and digital criminology. It then discusses key debates of interest, honing in on questions related to knowledge and how digital devices mediate knowledge production for the purpose of border control. It last explores foundational questions of ontology and epistemology about how digital criminology itself can produce knowledge about borders and border control.

Keywords: borders, knowledge, digital, mediation

Introduction

The border is most commonly understood as the territorial demarcation between two states. This spontaneous understanding is partly misleading, but forms a useful starting point for the discussion to follow. The border is not naturally given, but is a historical, social institution and construction symbolically and physically manifested by infrastructure ranging from simple roadside signage to barriers, checkpoints or fences, landscape features given meaning as territorial edges (e.g., mountains, rivers), as well as procedures and rituals (checks on documents, body searches, vehicle inspections, and so on). All of these are integral to the practices, that is the meaningful repeated performances, grouped under the heading of border control. In the last thirty years or so, the practices manifesting the border and enacting border control have noticeably transformed through the deployment of a wide range of digital devices, including (non-exhaustively) large-scale information systems and databases, biometric sensors, computer interfaces as well as algorithms and automated software processes (see Automation by Mann). These transformations have largely occurred in relation to security as well as law and immigration enforcement concerns, and border control is accordingly an important area of interest in the field of digital criminology.

In order to account for this interest, the chapter first discusses the relation between digital criminology and border criminology, that is the domain of criminological scholarship that has made borders and border control its main concern. It then probes key debates of interest for digital criminology. These involve questions of knowledge, concerning first how digital devices mediate knowledge production for the purpose of border control, and second more foundational questions of ontology and epistemology about how digital criminology itself can produce knowledge about border control.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-015

Border criminology

To discuss how borders and border control matter for digital criminology, we should first examine how criminological scholarship in general has taken notice of this issue. In an early discussion, Bowling (1990: 483) noted that a "key issue on the European criminological agenda is the control of intra-European movement as border controls begin to disappear in the early 1990s," signposting interrelated concerns with "the movement of drugs, 'criminals', 'terrorists', 'hooligans', 'economic migrants', and refugees," and locating criminological concerns with the border and border control at the intersection between migration, crime, and security enforcement. Subsequent work on transnational policing (e. g., Sheptycki, 1995) highlighted more systematically how borders and border control processes were becoming key sites for the transformation of law enforcement organizations and the criminalization of transnational movements of goods, money, and persons (see Financial Crime and Surveillance by Amicelle). Since the 2000s, questions related to borders and border control have been meaningfully attended to by the domain of border criminology (Bosworth, 2008), also known as the criminology of mobility (Pickering et al., 2015).

Put succinctly, border criminology interweaves three major 'theses' (Weber and McCulloch, 2019). The 'crimmigration thesis' (Stumpf, 2006) holds that criminal and immigration law increasingly converge in terms of substance, enforcement, and procedure. The key condition of possibility for this convergence is that both criminal and immigration law "act as gatekeepers of membership [...], determining whether an individual should be included in or excluded from our society" (Weber and McCulloch, 2019: 396 – 397), the former implicitly (through incarceration and deprivation of civic rights), the latter explicitly (through immigration detention and removal). Initially US-specific, the crimmigration thesis highlights two key notions of border criminology in general. First, the border and border control are not exclusively about controlling access to the territory of the state, but about determining whether and how one belongs to a particular group, society, or political community. They concern social inclusion, exclusion, and sorting. Border control accordingly takes place in multiple sites, not just along the territorial edges of the state, and through manifold practices (for early takes, see e. g., Pickering and Weber, 2013 on Australia; Weber and Bowling, 2004, on Europe and the United Kingdom). The second overarching thesis found in border criminology is that of "governing through immigration [or border] control" (Bosworth, 2008). Because it polices belonging, border control constitutes a governance practice beyond matters of territorial access and affect foreigners and nationals of a given state alike. The third thesis is the pre-emptive or pre-crime border thesis, which concerns more specifically the 'when' of border control. Through efforts to make sense of the designation of border control as a critical (national) security function, particularly following the attacks of 11 September 2001 on US soil (Weber and McCulloch, 2019: 504), border criminology argues that the border is not only manifested at points of entry into state territory. Border control can and does occur both before a

person has entered said territory, but also before any immigration or criminal offence has been committed.

How borders and border control matter to digital criminology should be understood in relation to the main theses and themes of border criminology. The latter have from an early stage manifested an interest in the technological features of border control. In the collection edited by Pickering and Weber (2006), the focus is mostly on how increasingly automated "means of social control" (Pickering and Weber 2006: 9) relying on electronic data, such as biometric verification and identification, are effecting patterns of inclusion and exclusion among people on the move, facilitating the travels of some while relegating others to the most dangerous and difficult journeys. A decisive contribution linking the themes of border criminology and criminological preoccupations with digital devices is Katja Franko Aas' (2011). Drawing deeply from interdisciplinary references in critical approaches to security and surveillance studies, Aas foregrounds the growing number and scope of databases involved in European Union (EU) border governance and the way in which these databases collect and circulate the personal data of non-EU/third-country nationals for wide-ranging purposes encompassing and articulating border and migration enforcement with law enforcement and security rationales. She further points to how the establishment of such databases by public authorities combines with the creation of commercial databases under private authority that enable individuals to fast-track entry checks in exchange for a fee and personal data, including biometrics, and of public–private arrangements such as that involved in the processing of Passenger Name Record (PNR) data (e.g., Jeandesboz, 2021; Glouftsios and Leese, 2023).

In so doing, Aas highlights a key insight about border control that should matter to digital criminology, namely that the digital processing of data she identifies aims both at 'gate closing' and 'gate opening' for border crossers. Closings and openings operate along citizenship and nationality lines but not only, as "[t]he privilege of high mobility is not reserved only for EU citizens and, importantly, it is not a privilege enjoyed by all EU citizens. The purpose [...] is to carve out from the long lists of third country nationals the ones which are trustworthy," that is "bona fide" (Aas, 2011: 338). This speaks to a longstanding argument of surveillance studies, namely that the purpose of surveillance is "social sorting" (Lyon, 2003). Understanding border control as sorting, as producing conditional, multi-speed inclusion alongside exclusion, is possibly the most widely shared view among border and digital criminology, with the latter having explored how digital devices make a difference in effecting inclusion and exclusion. The 'how' of border control is one area where ongoing discussions within digital criminology are currently taking place.

How digital devices do mediate border control

Asking how digital devices mediate border control as a practice of social sorting raises the question of knowledge. How do persons crossing international borders become

known for the purpose of sorting? How and to what extent does digital mediation matter in this respect? At stake here is an issue that traverses the criminological literature on policing at large, namely understanding and characterizing how policing bodies, including border and immigration enforcement organizations, generate knowledge and make it actionable.

Discussions in the literature on digital mediation and border control have focused here on the relation between digitally mediated and 'traditional' or 'low-tech' modes of knowledge production (see Low-Tech by Vestad). Some scholars tend to define contemporary borders and border control principally by reference to the technologies they manifest through and their stated novelty, as 'digital' or 'technological' borders (e.g., Broeders, 2007; Dijstelbloem and Meijer, 2011), with the implication that digitally-mediated knowledge production replaces other, pre-existing practices. By contrast and to use an exemplar study, in their analysis of the production and use of Passenger Name Record (PNR) data, which is data on travelers generated for reservation and operational purposes by airlines and transmitted to border enforcement authorities for security purposes, Glouftsios and Leese (2023) characterize border control in terms of 'epistemic fusion.' Digitally-mediated knowledge production might *displace* but does not *replace* other forms of knowledge, including practices of intelligence-gathering or criminal investigation. Digitally-mediated knowledge produced about border crossers "becomes contextualized and resonates with more traditional forms of security work" (Glouftsios and Leese, 2023: 132; for policing at large Kaufmann et al., 2019, Egbert and Leese, 2021). The task ahead for digital criminology research on border control, then, is to probe the friction and fusion of different modes of knowing for the purpose of sorting international mobilities rather than assuming that digitally-mediated knowledge production has replaced other modes.

A second related discussion found in the literature examines the heterogeneity of digitally-mediated knowledge production for border control. The literature generally shares the view that digitally-mediated border control involves the reliance on profiling techniques and has become a terrain for the deployment of algorithmic knowledge production (see Algorithm by Leese). Profiling here consists in sorting border crossers by matching data held about them with certain characteristics or sets of characteristics that are associated with illicit conducts. These characteristics can be derived from clues or evidence gathered as part of an investigation (e.g., a travel document number, a set of fingerprints, a credit card number). They can alternatively be derived from practical knowledge or assumptions about a certain kind of illicit conduct (e.g., travel itineraries used for the purpose of trafficking people or goods, nationalities, age, or gender, and usually a combination of those). In this second case, profiling is "enacted in a confirmatory or hypothesis-testing way to explore whether certain patterns of characteristics are represented in the analyzed population data and if so, to put the identified individuals under scrutiny" (Leese, 2014: 498). Both evidence-based and hypothesis-based profiling are 'rules-based,' in the sense that they involve the prior establishment of a fixed set of rules to sort out data available to border enforcement authorities, and are deployed through rules-based algorithms. As Amoore (2021: 4) illustrates,

"a rules-based algorithm for calculating border risk might have arranged variables such as IF nationality X AND travel Y THEN high-risk ELSE low risk." Hers and other contributions, however, also find that a third kind of profiling is growingly manifesting in border control based on 'features' of the data generated about border crossers. Driven by the deployment of machine-learning models, 'feature-based' algorithms no longer rely on pre-established rules but on attributes inferred from the data examples fed into models. Put differently, rules that result in the sorting of border crossers in feature-based profiling are machine- rather than human-designed. The conclusion should however not be that 'feature-based' profiling is a more technologically advanced form of digital mediation or that it is replacing rules-based profiling (Glouftsios and Leese, 2023: 138). What these findings indicate, rather, is that digital mediation is heterogeneous and that there is not a single form of digitally-mediated knowledge production for border control.

Border control as a sociotechnical setting

Discussions on how digital devices mediate border control branch out into two additional sets of foundational ontological and epistemological considerations.

The ontological issue involves characterizing the kinds of entities that digital criminology is confronted with when dealing with contemporary border control. Should these entities be considered as wholly or predominantly social, or technological? Given the apparent extent in the deployment of digital devices, one could make the case that human intervention and agency is increasingly limited when it comes to sorting international mobilities, and that interaction between (non-human) devices increasingly determines sorting at the border (see Categorization and Sorting by Franko). However, most of the literature emphasizes the importance of moving beyond instrumental accounts of digital technology while warning against the shortcomings of deterministic accounts. In instrumental accounts, digital devices do not ultimately matter to the way borders are controlled, because they are fundamentally instruments of human purposes and plans, and merely participate in the implementation of the latter (Singler, 2023: 16). Digital devices are deployed to solve specific problems, in our case, to sort 'bona fide'/trustworthy from 'mala fide'/untrustworthy border crossers, and do so more or less according to (human) plan. By contrast, research findings in the literature dedicated to digitally-mediated border control show how the deployment and operation of digital devices and their properties can alter human purpose and plans through feedback effects or the generation of unintended consequences. Writing about one of the largest European information systems dedicated to border control among other purposes, the Schengen Information System (SIS), Bellanova and Glouftsios (2022: 170) highlight for instance the "flickering foundations" of digitally-mediated borders, whereby "data infrastructures have far-reaching effects while being constantly subject to errors and malfunctioning." Holding millions of records of non-EU citizens deemed inadmissible by EU countries, the SIS is simultaneously a powerful and fragile border

control device, a fickle data infrastructure whose deployment has had the unintended consequence of spawning an entire domain of activity dedicated to its constant maintenance and repair. Acknowledging the 'agentic properties' of digital devices, that is their capacity to affect, deflect, or redirect the purposes and plans underpinning their deployment should not, however, tip the analytical scales in the opposite direction by "representing the political effects of the system as entirely determined by the technical tools themselves" (Singler, 2023: 17). Digital devices become *entangled* with organizational and social dynamics within border policing organizations, rather than the determinant factor in these dynamics. In her study of frontline staff at the Canada Border Service Agency, for example, Côté-Boucher (2018) outlines how the way in which border guards make sense of their work in the uncertain context brought about by changing policy expectations, organizational standards, and technologies, is affected by professional dispositions and socialization, which they express in her case through "generational talk." "Officers' accounts", she writes, show that border control "evolves in a more multifaceted temporal world than previously thought – from nostalgia for a simpler past and accusations of anachronism to overconfidence in technologies as tools for the future" (Côté-Boucher 2018: 165). What seems to be a point of convergence in the literature, then, is that we simultaneously *populate* analytical accounts with non-human devices and their agentic properties and *people* these accounts with social dynamics in order to avoid considering the emergence of digitally-mediated border control as a seamless, frictionless process.

Understanding border control as a sociotechnical setting is also and lastly the basis for an ongoing epistemological debate in the literature on digital mediation and border control. As noted among border criminology's important voices, criminology at large has, along with other domains of scholarship in the social sciences and beyond, contributed to constitute global distinctions between 'North' and 'South' by, among other features, granting analytical prominence to Anglo-American contexts and aligning itself with Western authorities' power to make people and things 'illegal' (Aas, 2012). It is tempting to consider that digital border control devices, the purposes and plans attached to them are shaped in 'high-tech,' Western contexts and subsequently transferred elsewhere. Such a view, however, is factually questionable and analytically problematic. Frowd (2020: 148) finds for instance that what he calls the "biometric ideal" in border control, that is the "set of knowledge claims about the effectiveness and symbolism of biometrics," takes shape through transnational interactions between international and West African border control experts and practitioners. However, patterns of "emulation," which he considers the key mechanism accounting for the circulation of ideas, practices, and devices of border control, are also a result of "African agency." Deploying biometric identification as a digitally-mediated practice of border control is considered by West African practitioners as a "means of ensuring conformity to what are 'global' standards," but are also "relative to the states that are considered leaders in the area of border security" in the region, such as Mauritania in the case of Senegal (Frowd, 2020: 152).

Beyond factual accuracy, however, the challenge facing digital criminology when dealing with borders and border control involves confronting "epistemological bordering, through which the difference between the knowledge produced from the North and the South is maintained and the latter is designated as the 'other'" (Mehta, 2023: 2). This includes questioning, rather than starting from, ready-made demarcations between 'high-tech' and 'low-tech' sociotechnical settings in the field of border control, linear stories about the 'digitization' of border enforcement worldwide, as well as the silencing of historical experiences and subjectivities through abrupt categorizations such as that between 'bona' and 'mala fide' border crossers.

Main takeaways

In discussing how borders and border control matter for digital criminology, the chapter has outlined the following key points:

- While we tend to think about border control as the enforcement of territorial demarcations between states, border criminology shows that it stands for practices of sorting between wanted and unwanted ('bona/mala fide') border crossers, and as such produces conditional, multi-speed international mobilities in addition to preventing some altogether.
- Debates about digitally-mediated border control have involved in particular examining knowledge production for the purpose of sorting border crossers. Digital mediation is heterogeneous and characterized by processes of fusion and friction between different sources of data and modes of knowledge.
- Studying border control raises ontological and epistemological questions for digital criminology. Ontologically, the task requires digital criminology to both populate its analytical accounts with non-human devices and their agentic properties and people these accounts with social dynamics. Epistemologically, the task demands that digital criminology meaningfully attends to the Eurocentric features of criminological knowledge production and confronts epistemological bordering.

Suggested reading

Aas, K. F. (2011). 'Crimmigrant' bodies and bona fide travelers: Surveillance, citizenship and global governance. *Theoretical Criminology*, 15(3), 331 – 346.

Côté-Boucher, K. (2020). *Border Frictions: Gender, Generation and Technology on the Frontline.* London: Taylor & Francis.

Pickering, S., & Weber, L. (eds.) (2006). *Borders, Mobility and Technologies of Control.* Dordrecht: Springer.

References

Aas, K. F. (2011). 'Crimmigrant' bodies and bona fide travelers: Surveillance, citizenship and global governance. *Theoretical Criminology*, 15(3), 331 – 346.

Aas, K. F. (2012). 'The Earth is one but the world is not': Criminological theory and its geopolitical divisions. *Theoretical Criminology*, 16(1), 5 – 20.

Amoore, L. (2021). The deep border, *Political Geography*, doi: 10.1016/j.polgeo.2021.102547

Bellanova, R., & Glouftsios, G. (2022). Controlling the Schengen Information System (SIS II): The infrastructural politics of fragility and maintenance. *Geopolitics*, *27*(1), 160 – 184.

Bosworth, M., & Guild, M. (2008). Governing through migration control: Security and citizenship in Britain. *The British Journal of Criminology*, *48*(6), 703 – 719.

Bosworth, M. (2017). Border criminology and the changing naure of penal power. In A. Liebling, S. Maruna, & L. McAra (eds.), *Oxford Handbook of Criminology, 6th Edition* (pp. 373 – 390). Oxford: Oxford University Press.

Bowling, B. (1990). Conceptual and methodological problems in measuring 'race' differences in delinquency. *British Journal of Criminology*, 30(4), 483 – 492.

Broeders, D. (2007). The new digital borders of Europe: EU databases and the surveillance of irregular migrants. *International Sociology*, 22(1), 71 – 92.

Côté-Boucher, K. (2018) Of "old" and "new" ways: Generations, border control and the temporality of security. *Theoretical Criminology*, 22(2), 149 – 168.

Dijstebloem, H., & Meijer, A. (eds.) (2011). *Migration and the New Technological Borders of Europe.* Basingstoke: Palgrave Macmillan.

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London: Routledge.

Frowd, P. M. (2020). *Security at the Borders: Transnational Practices and Technologies in West Africa.* Cambridge: Cambridge University Press

Glouftsios, G., & Leese, M. (2023). Epistemic fusion: Passenger Information Units and the making of international security. *Review of International Studies*, 49(1), 125 – 142.

Jeandesboz, J. (2021). 'Ceci n'est pas un contrôle': PNR Data Processing and the reshaping of borderless travel in the Schengen Area. *European Journal of Migration and Law*, 23(4), 431 – 456.

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *British Journal of Criminology*, 59(3), 674 – 692.

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue*, 45(5), 494 – 511.

Lyon, D. (2003). Surveillance as social sorting: Computer codes and mobile bodies. In D. Lyon (ed.), *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination* (pp. 13 – 30). New York: Routledge.

Mehta, R. (2023). A Southern feminist approach to the criminology of mobility. *International Journal for Crime, Justice and Social Democracy*, 12(2), 1 – 12.

Pickering, S., Bosworth, M., & Franko Aas, K. (2015). The criminology of mobility. In S. Pickering & J. Ham (eds.), *Routledge Handbook on Crime and International Migration* (pp. 382 – 395). London: Routledge.

Pickering, S., & Weber, L. (eds.) (2006). *Borders, Mobility and Technologies of Control.* Dordrecht: Springer.

Pickering, S. & Weber, L. (2013). Policing transversal borders. In K. F. Aas & M. Bosworth (eds.), *The Borders of Punishment* (pp. 93 – 110). Oxford: Oxford University Press.

Sheptycki, J. (1995). Transnational policing and the makings of a post-modern state. *British Journal of Criminology*, 35(4), 6103 – 6135.

Singler, S. (2023) Performativity, pragmatism and border control technologies: Democratising the ontologies of border criminology. *International Journal for Crime, Justice and Social Democracy*, 12(2), 13 – 24.

Stumpf, J. (2006). The crimmigration crisis: Immigrants, crime, and sovereign power. *American University Law Review*, 56(2), 367 – 419.

Weber, L., & Bowling, B. (2004). Policing migration: A framework for investigating the regulation of global mobility. *Policing and Society*, 14(3), 195 – 212.

Weber, L., & McCulloch. J. (2019). Penal power and border control: Which thesis? Sovereignty, governmentality, or the pre-emptive state? *Punishment & Society*, 21(4), 496 – 514.

Katja Franko

15 Categorization and sorting

Abstract: This contribution examines the criminological and social relevance of categorization and social sorting, its historic roots, present practices, and implications for the future. The chapter examines the transformation of social sorting from being the corner stone of statecraft to becoming a central feature of contemporary capitalism. These developments have profound implications for criminological understanding of surveillance, social control, and crime control, which are increasingly not only digitalized but also privatized and commercialized.

Keywords: social sorting, capitalism, commercialization, legibility

In August 2022, Reuters reported about a Turkish sports presenter Sinem Okten who was surprised to see her visa application to Europe's Schengen area rejected twice, despite frequent previous job visits to the continent.[1] A considerable proportion of applicants from Turkey are denied visas. To process its visa applications, the European Union operates one of the largest biometric databases in the world. The capacity of the VIS database has increased in recent years to 100 million records and 85 million biometric matching records, which allow border authorities to identify and verify third-country nationals who travel to the EU.[2] In the post-9/11 world, biometrics have become one of the preferred methods of social control used by state authorities concerned about security, terrorism, organized crime, and unauthorized migration.

Modern states create massive bureaucratic records, which include increasingly sophisticated information, to facilitate their decision-making. The EU is far from unique in this respect as countries such as China and India have reportedly even larger biometric databases. As James C. Scott (1998: 65) shows in his seminal work, *Seeing Like a State*, the creation of legible people—"of fixing an individual's identity and linking him or her to a kin group"—has been a necessary precondition of modern statecraft (see also Lyon, 2009 on ID cards and Caplan and Torpey, 2001 on the creation of the passport). Technologies such as passports and biometrics attempt to make vast numbers of the world population legible not only to states they are citizens of, but also to other states (Franko, 2020).

However, as the above example of the Turkish journalist reveals, these bureaucratic processes entail not only categorization, but also carry an inherent possibility of social sorting and, ultimately, social exclusion. As state functionaries classify people into appropriate categories, based on their fingerprints and other information, they are also

1 https://www.reuters.com/world/turks-frustrated-by-deliberate-increase-number-european-visa-rejections-2022-08-26/.

2 https://www.eulisa.europa.eu/Publications/Reports/2021%20VIS%20Report.pdf.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-016

sorting out undesirable visa applicants. The EU's ambition in the past two decades has been to make the information on visa applicants and residence permit holders interoperable with police registers in order to allow border guards, police officers, and immigration officials "to have more complete information on security threats as well as seamless access to information."[3] Notions of security, risk, threat, and crime are therefore central in defining and structuring the contemporary activities of categorization and social sorting.

This contribution examines the criminological and social relevance of categorization and social sorting, its historic roots, present practices, and implications for the future. Categories are, as Bowker and Star (2000: 196) observe, something that people often take for granted, often forgetting that classification systems are sites of negotiation and political and social struggles. Moreover, it is important to keep in mind that putting people into categories according to various social parameters, and treating them differently depending on these categories, is not only something that state agencies have done for centuries. It is also something that is done by private companies and other actors, increasingly for commercial purposes. As consumers and potential consumers, bank customers, students, internet and social media users, information about our activities and digital movements is analyzed and sorted for its commercial relevance. The chapter therefore also examines the transformation of social sorting from being the corner stone of statecraft to becoming a central feature of contemporary capitalism. These developments have profound implications for criminological understanding of surveillance, social control, and crime control, which are increasingly not only digitalized but also privatized and commercialized (see Privatization by Lomell).

(Digital) surveillance and social sorting

Creation of categories is, as Scott's work shows, one of the essential aspects of modern statehood. It is not only people that are made legible by being put into *categories*, but also space and nature under state jurisdiction are transformed into "closed systems that offer no surprises and that can best be observed and controlled" (Scott, 1998: 82). Historically, these ambitions became clearly developed during the 19th century and have given rise to increasingly sophisticated census making, collection of statistical information about numerous aspects of economic, demographical, and other social activities (for a history of statistics see Hacking, 2015). Michel Foucault influentially described these developments as a form of biopolitics, in which "government has as its purpose not the act of government itself, but the welfare of the population, the improvement of its condition, the increase of its wealth, longevity, health, etc" (Foucault, 1991: 100). Throughout the 19th and 20th centuries, the bio-politically oriented state

3 https://www.eulisa.europa.eu/Publications/Reports/2021%20VIS%20Report.pdf.

showed a growing thirst for knowledge in order to achieve greater productivity and welfare of the nation, moral conformity, and compliance of the population. These state ambitions also profoundly shaped the nature of policing and prisons, which became central links in state projects of incorporating all sections of society into an ordered, surveilled, and productive citizenry.

Due to these processes of knowledge collection, states have not only amassed large amounts of data about the population but have also created categories into which people are placed and through which we still understand ourselves today. As Hacking (2015: 66) observes: "Enumeration demands *kinds* of things or people to count. Counting is hungry for categories. Many of the categories we now use to describe people are byproducts of the needs of enumeration" (italics original). Statistics "as a moral science of the state" laid the ground for, and has been a driving force behind the rise of calculating machine technologies that came to prominence in the 20th century and so strongly define contemporary modes of governance (Bigo et al., 2019: 3). The historic origins of state categorization and data collection show how knowledge and power are intimately connected, which prompted Foucault (1982) to coin the famous power/knowledge nexus. The will to knowledge and the will to power are thus, as Bigo et al. (2019: 6) point out, "two aspects of how we conduct ourselves and the conduct of others."

Drawing on Foucault's work, particularly his concept of the panopticon, a large body of surveillance and criminological scholarship has in recent decades explored the connections between state and its increasing capabilities to automatically collect data about our daily lives. A central point made by this scholarship is that surveillance practices are intrinsically connected to, and inseparable from, practices of *social sorting*. In his large body of work on the subject, David Lyon (2003) defines social sorting as practices of placing people into social classes and categories, which then enable those making the categories to distinguish between desirable and undesirable populations. The concept of social sorting places the issue of automated data collection "in the social and not just the individual realm – which 'privacy' concerns all too often tend to do" (Lyon, 2003: 13). It highlights the classifying drive of contemporary surveillance, as well as defuses some of its sinister (and conspiratorial) aspects and plants the issue firmly in the domain of everyday life. As Lyon (2003: 13) points out: "Human life would be unthinkable without social and personal categorization, yet today surveillance not only rationalizes but also automates the process."

In the past three decades, the rapid development of large-scale data processing capabilities and the political climate of the so-called *war on terror* have made the question of social sorting one of the most pressing ethical, political, and legal dilemmas of our time (Bigo et al., 2019). In the aftermath of the September 11 attacks, and the subsequent terror attacks in several cities across the world, practices of categorizing people and assessing their potential risk have proliferated (Franko, 2020). Surveillance and social sorting practices proactively exclude some social groups whose future behavior is considered undesirable (see inter alia Amoore and De Goede, 2008; Selod, 2018). Although the proliferation of digitally enabled suspicion seems to affect most citizens it is

important to keep in mind, however, that not everyone is affected in the same way. A growing body of surveillance scholarship has brought to attention how the nature of surveillance and social sorting is shaped by class, race, and gender (see inter alia Browne, 2015; Monahan, 2022). Just as most criminologists would be aware of Foucault's (1977) argument that surveillance and disciplining technologies historically targeted a particular social figure—broadly defined as the 'delinquent'—contemporary practices of social sorting disproportionately focus on some disadvantaged groups, such as Muslim men (Selod, 2018) or particular categories of immigrants (Franko, 2020).

The case of the Turkish journalist denied entry into the EU is, therefore, part of a broader pattern where suspicious groups of racialized, poor, and less affluent travelers find themselves under intensified suspicion and denied entry. For affluent business travelers, tourists, and citizens of wealthy countries, on the other hand, fingerprint technologies and dedicated databases may mean that their movements across borders can become faster and easier due to frequent traveler programs and biometric passports (Aas, 2011). While unwanted migrants get sorted out, these groups of 'bona fide' travelers get 'sorted in.'

Surveillance capitalism and commercial sorting

The example of frequent traveler programs also serves as a reminder that the rise of big data surveillance has not only dramatically expanded state capabilities for collection of data but has also enabled private and commercial actors to operate large databases (see Big Data by Završnik). The state thus no longer has a monopoly, and may not even be the main actor, when it comes to collection of large amounts of data and consequent social sorting of individuals (Bigo et al., 2019). Today, commercial actors such as Google and Facebook command surveillance capabilities that match and exceed those of most states. While the big data surveillance capabilities can be co-opted into and amplify state's law enforcement objectives, they also profoundly change the nature of contemporary law enforcement. In the past two decades, there has been a trend towards predictive policing where analytical techniques are used to make statistical predictions about where and when potential crimes might occur and who might be the perpetrators and the victims (Brayne, 2020).

The commercial aspects of surveillance were brought to attention already in 1993 by Oscar Gandy's seminal study *The Panoptic Sort.* The book was highly critical of the failure of scholars and political activists to pay sufficient attention to the threats to privacy posed by commercial firms. In the following decade, the field of surveillance studies and criminologists directed growing attention to social sorting and exclusion conducted by commercial actors and in commercial spaces, particularly those using CCTV surveillance (Norris, 2012). McCahill and Finn (2014) suggested that the use of surveillance to proactively exclude some social groups whose future behavior is considered undesirable can be described as ban-opticon (see also Bigo, 2006); it functions

in a way as a membrane that includes some and excludes others. Commercial spaces, like shopping malls, are an ideal example of this: 'Flawed consumers' are pushed away, while ideal consumers are kept in, and CCTV cameras are mainly directed towards those who do not belong.

Although important, these early studies of commercial sorting nevertheless focused mainly on the *physical* presence of undesirable individuals and groups, which pale in comparison to the extent of contemporary *digital* sorting by commercial actors. The internet has dramatically enhanced consumer profiling. Many websites routinely install tracking technologies on computers of their users and create databases of consumer profiles (Andrejevic, 2007; see Databases by Bellanova). One only has to think of how a routine purchase in an online shop usually results in a series of 'personalized' recommendations and commercial email offers. Data gathering is a vital aspect of the growing e-commerce, work-related surveillance and even politics—a development that has been captured by Shoshana Zuboff's (2019) influential term (and eponymous book) surveillance capitalism. By tracking information about our purchases, browsing histories, movements, statements, and other aspects of private life, companies such as Google and Facebook are creating and commercially exploiting the "'behavioural surplus'" of our actions and not only turning it into profit, but more importantly into a new "means of behavioural modification" (Zuboff, 2019). While, traditionally, the state and its institutions, such as prisons, schools, and psychiatric institutions, possessed the most powerful means of classification and behavioral modification, this may no longer be the case.

Conclusion

In 2018, critical observers, political analysts, and ordinary citizens alike were astounded by the revelations of a whistleblower, Christopher Wiley, that data of up to 87 million Facebook users was improperly shared with the political consultancy Cambridge Analytica without proper consent to allegedly influence the 2016 US presidential election.[4] Although commercial data harvesting was becoming common knowledge, the extent of the collusion of commercial actors and political profiling was nevertheless a surprise to many and prompted government hearings in the US, UK, and Canada (Lyon, 2019: 64). The scandal showed that digital data have both a commercial and a political dimension, and that is all but impossible to separate the two. The integration or merging of state and commercial surveillance capabilities, always a latent possibility and a frequent dystopian vision, is becoming a more palpable reality as voting consumer and social media profiles can be combined for analytical purposes of those in power. Systematic harvesting of commercial and personal data, their categorization, and sorting thus form the basis for subsequent profiling of individuals and create the potential

4 https://www.bbc.com/news/technology-43649018.

for reinforcement of so-called echo chambers. An echo chamber "can act as a mechanism to reinforce an existing opinion within a group and, as a result, move the entire group toward more extreme positions" (Cinelli et al., 2021: 1), a development that is a central component of on-line radicalism and political polarization.

However, Cambridge Analytica scandal also revealed that there are possibilities for resistance and that critical journalism, consumer, and citizen awareness as well as political oversight still carry considerable weight. In December 2022, Facebook owner Meta agreed to pay a $725 m (£600 m) settlement, the largest in a US data privacy class action. The company had also wowed to 'revamp' its approach to privacy. The sum was nevertheless minor compared to the company's profits and commercial investment.[5] The contestations surrounding the Cambridge Analytica scandal illustrate Bowker and Star's (2000) point that classification practices are not something that should be taken for granted but are sites of intense political and social struggles. They, therefore, deserve continued scholarly interest and careful scrutiny.

Suggested reading

Aas, K. F. (2011). 'Crimmigrant' bodies and bona fide travelers: Surveillance, citizenship and global governance. *Theoretical Criminology*, 15(3), 331 – 346.

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing.* New York: Oxford University Press, USA.

Lyon. D. (ed.) (2003). *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination.* London and New York: Routledge.

Monahan, T. (2022). *Crisis Vision: Race and the Cultural Production of Surveillance.* Durham, NC: Duke University Press.

Scott, J. C. (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition have Failed.* New Haven: Yale University Press.

References

Aas, K. F. (2011). 'Crimmigrant' bodies and bona fide travelers: Surveillance, citizenship and global governance. *Theoretical Criminology*, 15(3), 331 – 346.

Amoore, L., & De Goede, M. (eds.). (2008). *Risk and the War on Terror.* London: Routledge.

Andrejevic, M. (2007). *iSpy: Surveillance and Power in the Interactive Era.* Lawrence, KS: University Press of Kansas.

Bigo, D. (2006). Security, exception, ban and surveillance. In D. Lyon (ed.), *Theorizing Surveillance: The Panopticon and Beyond* (pp. 46 – 68). London: Willan.

Bigo, D., Isin, E., & Ruppert, E. (2019). Data politics. In D. Bigo, E. Isin, & E. Ruppert (eds.), *Data Politics: Worlds, Subjects, Rights* (pp. 1 – 17). London: Routledge.

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and its Consequences.* Cambridge, MA: MIT Press.

5 https://www.bbc.com/news/technology-64075067.

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing.* New York: Oxford University Press, USA.

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness.* Durham, NC: Duke University Press.

Caplan, J., & Torpey, J. (2001). *Documenting Individual Identity: The Development of State Practices in the Modern World.* Princeton, NJ: Princeton University Press.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118.

Foucault, M. (1977). *Discipline and Punish* (Alan Sheridan, trans.). New York: Pantheon.

Foucault, M. (1982). The subject and power. *Critical Inquiry*, 8(4), 777 – 795.

Foucault, M. (1991). *The Foucault Effect: Studies in Governmentality.* Chicago: University of Chicago Press.

Franko, K. (2020). *The Crimmigrant Other: Migration and Penal Power.* Abingdon: Routledge.

Gandy, O. H. (1993/2021). *The Panoptic Sort: A Political Economy of Personal Information.* Oxford: Oxford University Press.

Hacking, I. (2015). Biopower and the avalanche of printed numbers. In V. W. Cisney & N. Morar (eds.), *Biopower: Foucault and Beyond* (pp. 65 – 80). Chicago: University of Chicago Press.

Lyon, D. (2003). Surveillance as social sorting: Computer codes and mobile bodies. In D. Lyon (ed.), *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination* (pp. 13 – 30). London and New York: Routledge.

Lyon, D. (2009). *Identifying Citizens: ID Cards as Surveillance.* Malden, MA and Cambridge: Polity Press.

Lyon, D. (2019). Surveillance capitalism, surveillance culture and data politics. In D. Bigo, E. Isin, & E. Ruppert (eds.), *Data Politics* (pp. 64 – 77). London: Routledge

McCahill, M., & Finn, R. L. (2014). *Surveillance, Capital and Resistance. Theorizing the Surveillance Subject.* Abingdon: Routledge.

Monahan, T. (2022). *Crisis Vision: Race and the Cultural Production of Surveillance.* Durham, NC: Duke University Press.

Norris, C. (2012). The success of failure: Accounting for the global growth of CCTV. In K. Ball, K. Haggerty, & D. Lyon (eds.), *Routledge Handbook of Surveillance Studies* (pp. 251 – 258). London: Routledge.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition have Failed.* New Haven: Yale University Press.

Selod, S. (2018). *Forever Suspect: Racialized Surveillance of Muslim Americans in the War on Terror.* New Brunswick, NJ: Rutgers University Press.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* London: Profile Books.

Stefano Mazzilli Daechsel

16 Computation

Abstract: This chapter provides an introduction to the computational logics and technologies that are increasingly (re)structuring our societies and the field of criminology. I start by providing a general overview of computation, computational theory, and the technological innovations that underpin contemporary general-purpose computers. I then critically examine ongoing efforts to integrate computational methods into criminological research under the banner of computational criminology.

Keywords: computation, computational theory, computational criminology

Introduction

If digital criminology means to take seriously the "technosocial nature of contemporary social and political life" (Powell et al., 2018: 4), the logics and material components of digital technologies ought to be treated as socially relevant forces. The ubiquity of computational devices in our societies dictates that digital criminologists should have at least a basic understanding of computation and computational theory in order to be attuned to their (re)structuring roles wherever computers are used, including in criminology itself. To this end, I start by providing a general overview of computation, computational theory, and the technological innovations that underpin contemporary general-purpose computers. I then pivot to critically examine ongoing efforts to integrate computational methods into criminological research under the banner of computational criminology.

Computation, computational theory and the general-purpose computer

At a basic level, computation refers to any process that acts on an input to produce an output. Most computation today is digital and electrical (think laptops and smartphones; see Digital by Wernimont) but computational technologies, such as the abacus,

Disclaimers and acknowledgements: Grant numbers and/or funding information: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 947681).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-017

have been around for millennia. Until about the mid-20th century, the most sophisticated computers were highly elaborate constellations of mechanical parts that could perform advanced calculations but were prone to deterioration and breakdown. Today, quantum and molecular computing are testing new models and materials for computation that could result in entirely new applications and capabilities.

Underpinning this great variety of computational technologies is computational theory, a field that has its origins in the discipline of mathematics. Just as computers, broadly defined, are thousands of years old, "the notions of computability and computable functions go back a long time" (Fernandez, 2009: 1), certainly as far back as the ancient Greeks and Egyptians. While figures such as Charles Babbage, Ada Lovelace, and George Boole are often cited as progenitors of modern computing, modern computational theory emerged from efforts in the 1930s by the likes of Alan Turing and Alonzo Church to respond to a series of mathematical problems first posed by David Hilbert at the International Congress of Mathematicians in 1900.

Turing's 1937 article, 'On Computable Numbers, with an Application to the *Entscheidungsproblem*,' describes an abstract computing machine that is composed of an infinite tape divided into squares, "each capable of bearing a 'symbol'" (Turing, 1937: 231), a head that can move along the tape from square to square, and a control unit that dictates how the head should behave given the content of the square it finds itself on. This *Turing machine*, as it is now known, provides the basic elements of the general-purpose computer—a memory unit (the tape), a central processing unit (the head), and a control unit—but Turing's immediate aim in positing this abstract machine was to determine whether Hilbert's *Entscheidungsproblem*, or decision problem, was a computable function. Turing asserts that a function is only computable if it allows the Turing machine to reach a final state, i. e., a final square on the tape, that provides the function's output or solution. The article concludes that Hilbert's decision problem is not a computable function because the Turing machine never reaches a final state but goes on moving from square to square on the tape forever.

Having a working understanding of Turing machines offers some important insights into the nature of computation and computational logics. Turing machines are abstract and thus not bounded by the strictures of the physical universe. Although many physical computers resemble Turing machines in their composition, even the most advanced supercomputers do not have an infinite memory unit, nor infinite energy and time. The abstract plane of computational theory often slams into the material and temporal constraints of reality; what is abstractly computable might not be feasibly computed by a physical machine. These limitations are especially important to keep in mind when computers are used to model and simulate technosocial realities that are highly complex and informationally dense.

To get from the Turing machine to the digital general-purpose computer, three other elements needed to fall into place. John von Neumann's 1945 *First Draft Report on the EDVAC* is considered the first description of a general-purpose electronic computer that can concretely implement all the components of a Turing machine. The *von Neumann architecture* detailed in the report is still dominant today in the world

of electronic computing. Information, which is physically instantiated in electrical circuits, passes between the memory unit and the central processing unit in accordance with the inputs provided by the user. Contemporary digital computers essentially embody this architecture.

The works of the mathematician and electrical engineer Claude Shannon were yet another major turning point. In 1937, Shannon wrote what is often considered the most influential master's thesis of all time entitled 'A Symbolic Analysis of Relay and Switching Circuits.' His thesis demonstrated an equivalence between electrical circuits and Boolean algebra, "a symbolic method of investigating logical relationships" (Shannon, 1938: 714) between two variables the result of which is either TRUE or FALSE. Given that this binary logic corresponds to electrical circuits that can be in one of two states, namely, ON or OFF, Shannon showed that electrical circuits could be designed to emulate specific logical relationships between two variables, such as 'AND,' 'NOT,' and 'OR.' His work led to the development of electronic logic gates that continue to form the underlying structure of all digital computers. Shannon's 1948 article, 'A Mathematical Theory of Communication,' is an early classic in the field of information theory and, among many other innovations, established the 'bit' (short for binary digit) as the unit for measuring information in computing and telecommunications.

The binary logic of electronic computing reduces all information to sequences of two discrete states in electrical circuits, ON and OFF, represented symbolically by the bits 1 and 0. On the one hand, the ability to encode virtually any kind of content in a sequence of bits is an incredibly powerful innovation that greatly facilitates the storage and transmission of digital information. On the other hand, converting non-digital information—a sound wave for example—into digital information comes with some potential downsides. As the continuous is made discrete, selection and flattening processes are at work that transform some of the original information and exclude the rest. In many circumstances, this trade-off is absolutely worthwhile, but it is crucial to understand that digital information does not capture the fullness of non-digital reality.

The third and final development to consider in the genesis of the general-purpose computer is the invention of the transistor in 1947 by John Bardeen, Walter Brattain, and William Shockley at the same Bell Labs where Shannon spent much of his career. Often considered "the fundamental building blocks of all modern electronic devices" (Konkoli et al., 2018: 156), transistors are semiconductor devices made of silicon (hence the name Silicon Valley) that can switch or amplify electrical signals. The decades-long trend of transistor miniaturization has meant that more and more transistors can be inserted on an integrated circuit, or microchip, thereby boosting a device's computing power without increasing its physical size. In 1965, the engineer and businessman Gordon Moore speculated that the number of transistors on a microchip would double every year, which he revised to every two years in 1975, a prediction that has largely held and earned the name 'Moore's law.' It is thanks to this process of transistor miniaturization that we are now able to walk around with computers in our pockets that are more powerful than supercomputers that once filled entire laboratories.

However, Moore's law will soon be made obsolete by the simple fact that transistors cannot be miniaturized past the point at which the classical laws of physics give way to quantum mechanics. This real limit on transistor miniaturization, coupled with the immense energy cost of conventional computing, has led to a lot of research on alternatives to silicon-based computing. Quantum computing and molecular computing have received the lion's share of interest and funding, but neither has proven itself capable of exhibiting the versatility and reliability of conventional computers. To date, it appears that these unconventional computing systems will supplement rather than replace conventional computers, finding their true utility in specific tasks.[1]

Criminology and computational power

Part of digital criminology's mandate is not only to understand the impact of computational logics and technologies on crime and crime control, but also on criminology itself. As a field, criminology has long embraced the use of computational methods to store and sort empirical data, such as crime statistics, survey findings, or other datasets (see Datafication by Chan). The larger these datasets have become, the more computer programs are called upon to perform analytical tasks "traditionally undertaken by social scientists" (Williams et al., 2017: 337). This has contributed to the emergence of what several criminologists have dubbed "computational criminology" (Berk, 2008; Williams et al., 2017; Campedelli, 2022; Steinmetz, 2023). Here, however, we must be careful to differentiate between two divergent approaches working under this banner.

On the one hand, there are researchers for whom computational criminology is an interdisciplinary methodology that engages with big data, particularly social media data (see Social Media by Twigt), to help address criminological questions (Williams and Burnap, 2016; Williams et al., 2017; Wiliams et al., 2020). This version of computational criminology sees potential in online data "to complement and augment conventional curated data" (Williams et al., 2017: 321). Computational methods are used to analyze and correlate online data with other datasets, all with the understanding that online data, particularly geolocated and time-stamped data, can provide new insights into long-standing criminological problems. As an example, Williams and Burnap (2016: 217) provide an "analysis of social media data using advanced computing techniques to answer a classic criminological question on social reactions to criminal events of national interest," which in their case was the Woolwich terrorist attack of 2013.

On the other hand, computational criminology is often associated with the development and application of computer simulations for criminological research (Berk, 2008; Malleson et al., 2010; Berk, 2013; Birks, 2018; Groff et al., 2019; Campedelli,

1 This swift overview of the theoretical and technological innovations that underpin contemporary computers should be coupled with the social history of computing and the computerization of society (see Mahoney, 2005, 2011; Campbell-Kelly and Aspray, 2004).

2022). Malleson et al. (2010), for instance, outline a computer model to simulate the occurrence of residential burglaries in different neighborhood constellations. The authors employ an agent-based model that simulates the behavior of artificial agents ('citizens' and 'potential burglars') that are assigned specific rules of interaction and certain drives, such as the need to sleep and the need to generate wealth (Malleson et al., 2010: 239). In this variation of computational criminology, abstraction and highly reductive reasoning are used to transform complex social environments and individuals into computable functions.

Advocates of computer simulations in criminology note that processes of abstraction are inherent to the formulation of any theory (Birks, 2018). Indeed, all theorists abstract the dynamics and variables they deem salient to a given problem from the complexity of social life. The difference, however, lies in the degrees and layers of abstraction needed to reduce social complexity to a computable model that simulates both agents and their environments. Whereas the first approach to computational criminology aims to develop methodological tools to better explore the significance of digital data that are already circulating and acting within society, the second form of computational criminology largely supplants the social with models and simulations that, no matter how sophisticated, are reductive abstractions that must comply with the logics and material constraints (time, energy, memory, computing power) of existing technologies. Even if one assumes "an infinite amount of computational power at [one's] disposal, it remains very challenging to try to accurately model a virtual society in all its facets and dimensions" (Campedelli, 2022: 63).

Conclusion

My aim in this chapter was to provide a general introduction to the computational logics and technologies that are increasingly (re)structuring our societies and the field of criminology itself. As digital criminologists, we should challenge strict binary oppositions between technology and society and advance a technosocial approach that treats technological logics and systems as social forces that must be understood on their own terms. Thus, here are the chapter's main takeaways:

- The history of computation reveals a process by which abstract machines, such as the Turing machine, and digital logics, such as Boolean algebra, gradually materialized into the general-purpose computers that we are all familiar with today. Unlike abstract machines, however, material computers are constrained by factors such as time, energy, memory, and computing power. Because of these constraints, computer scientists are exploring new computational models and materials, such as quantum computing and molecular computing.
- Computational logics are incredibly powerful and versatile when the information in question is digital (bits of 1s and 0s). However, translating non-digital information into digital information always involves processes of abstraction and simplification (see Translation by Wilson-Kovacs). The original signal is never fully cap-

tured by its digital counterpart. Digital criminologists should take account of these constraints when they study computers as knowledge-making devices or as a part of social practices.
– Criminologists have long used computers as part of their research, but there are ongoing efforts to integrate more sophisticated computational methods into criminology. The rise of computational criminology reflects this trend, although we must be attentive to the different approaches working under the same banner.
– While some computational criminologists are developing computational methods to integrate existing digital information, such as social media content, into criminological research, others are constructing computer models to simulate highly complex social dynamics, models that invariably reduce social reality to oversimplified (but computable) abstractions.

Suggested reading

Berk, R. (2013). Algorithmic criminology. *Security Informatics*, 2(5), 1 – 14.
Mahoney, M. S. (2005). The histories of computing(s). *Interdisciplinary Science Reviews*, 30(12), 119 – 135.
Steinmetz, K. (2023). The mundanity of computational criminology. In *Against Cybercrime* (pp. 155 – 187). London: Routledge.
Williams, M., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *British Journal of Criminology*, 57, 320 – 340.

References

Berk, R. (2008). How you can tell if the simulations in computational criminology are any good. *Journal of Experimental Criminology*, 4, 289 – 308.
Berk, R. (2013). Algorithmic criminology. *Security Informatics*, 2(5), 1 – 14.
Birks, D. (2018). Computer simulations: Agent-focused environmental criminology. In G. Bruinsma & S. Johnson (eds.), *The Oxford Handbook of Environmental Criminology* (pp. 311 – 340). Oxford: Oxford University Press.
Campbell-Kelly, M., & Aspray, W. (2004). *Computer: A History of the Information Machine*, Boulder, CO: Westview Press.
Campedelli, G. M. (2022). *Machine Learning for Criminology and Crime Research: At the Crossroads.* London: Routledge.
Fernandez, M. (2009). *Models of Computation: An Introduction to Computability Theory.* London: Springer.
Groff, E., Johnson, S., & Thornton, A. (2019). State of the art in agent-based modeling of urban crime: An overview. *Journal of Quantitative Criminology*, 35, 155 – 193.
Konkoli, Z., Stepney, S., Broersma, H., Dini, P., Nehaniv, C., & Nichele, S. (2018) Philosophy of computation. In S. Stepney, S. Rasmussen, & M. Amos (eds.), *Computational Matter* (pp. 153 – 184). London: Springer.
Mahoney, M. S. (2005). The histories of computing(s). *Interdisciplinary Science Reviews*, 30(12), 119 – 135.

Mahoney, M. S. (2011). *Histories of Computing.* Cambridge, MA: Harvard University Press.
Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems*, 34, 236 – 250.
Powell, A., Stratton, G., and Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society.* London: Routledge.
Shannon, C. (1938). A symbolic analysis of relay and switching circuits. *Transactions of the American Institute of Electrical Engineers*, 57(12), 713 – 723.
Shannon, C. (2001). A mathematical theory of communication *Mobile Computing and Communications Review,* 5(1), 3 – 55.
Steinmetz, K. (2023). *Against Cybercrime.* London: Routledge.
Turing, A. (1937). Computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(1), 230 – 265.
Von Neumann, J. (1993). First draft of a report on the EDVAC. *IEEE Annals of the History of Computing*, 15(4), 27 – 75.
Williams, M., & Burnap, P. (2016). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, 56, 211 – 238.
Williams, M., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *British Journal of Criminology*, 57, 320 – 340.
Williams, M., Burnap, P., Javed, A., Lui, H., & Ozalp, S. (2020). Hate in the machine: Anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *British Journal of Criminology*, 60, 93 – 117.

Thomas J. Holt and Karen M. Holt

17 Cybercrime

Abstract: This chapter provides an overview of the issue of cybercrime, or the use of the Internet and computer technology in order to offend. The definitional challenges associated with cybercrime are discussed, along with the various behaviors that have come to be defined as cybercrimes by criminologists.

Keywords: cybercrime, technology, online, Internet

Computers, mobile phones and the internet have revolutionized modern life. More than half of the world's population use the internet (Internet World Stats, 2020), and a majority of the populations of Western nations are accessing and utilizing this technology. This has transformed communication and relationships. Many now prefer to communicate via text message rather than voice phone calls (Zickuhr, 2011). Social media applications are now a critical resource for both personal communications and commerce (Perrin and Anderson, 2019). In addition, virtually all aspects of personal and business transactions are supported by the internet.

The benefits of technology are offset by the overwhelming opportunities they present for criminals who seek access to people, money, and digital information stored online (Holt and Bossler, 2015; Maimon and Louderbach, 2019). Computers and the internet have enabled and simplified traditional forms of crime, like acts of fraud (Button and Cross, 2017), stalking, and sexual offenses (Holt and Bossler, 2015). Technology has also created new forms of offending that are dependent on computers, such as hacking where individuals attempt to access computer networks without permission (see Hacking by Wall) (Steinmetz, 2016).

While much of the extant criminological literature has focused on the nature and qualities of online offenses, there is also an argument to be made that there has been an increase in the ability of the state and commercial institutions to centralize power and surveillance, which can cause more harm than online crimes themselves (Steinmetz, 2023; 4). The surveillance capabilities of the state have far-reaching implications, notably related to invasions of privacy, censorship, restriction of the internet, and the classification of citizens based on tracking their online movements. Thus, a critical approach to understanding cybercrime must examine not only the ways in which individuals can use technology for harm, but also how regimes can employ surveillance and use this technology under the guise of crime control (see Surveillance by Lyon).

Over the last two decades, criminologists have increasingly focused attention on the range of behaviors enabled by technology, which are commonly referred to as cybercrime (Leukfeldt and Holt, 2019; Payne and Hadzhidimova, 2020). Cybercrime is treated as an umbrella term which includes all offenses either occurring in or being wholly facilitated by an online environment. Thus, offenses occurring in physical spaces and involving minimal technological assistance would fall outside of the

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-018

scope of cybercrime. Here we will explore the meaning of this term, and the ways in which it has been operationalized by scholars over time. The range of behaviors that constitute cybercrimes will also be reviewed to provide a better appreciation for the scope of crimes occurring online.

Defining cybercrime

The current popularity and consistency in the use of the term cybercrime belies decades of debate among criminologists. In fact, some of the first pieces examining the misuse of technology in criminology can be found in the late 1980s, and used the term computer crime to reference the fact that a computer was involved to facilitate fraud at a Florida dog racing track (Hollinger and Lanza-Kaduce, 1988). At the time, the notion of the internet and cyberspace as we currently understand it was in its infancy.

As the World Wide Web became dominant in the 1990s and technology access flattened, criminologists began to consider the ways that the online environment may enable individuals to share information on offending that could be used on and offline (Mann and Sutton, 1998). David Wall (1998) argued that the growth of technology and online spaces created a phenomenon whereby not only were new offenses possible, but the environment in which they could occur did also not exist otherwise. For instance, the global access afforded by the internet makes it possible for offenders to identify victims in other countries with ease (Holt and Bossler 2015). Furthermore, computers act as a force multiplier for offenders, allowing them to affect hundreds of victims at the same time in ways that are not possible in physical space (Wall, 1998).

By contrast, Peter Grabosky (2001) argued that technology-enabled offenses were simply "old wine in new bottles," meaning they were traditional forms of crime occurring in a novel space using new tools. For instance, acts of fraud and theft could be performed in off-line spaces, though technology made it easier to do so online (Grabosky, 2001). The internet did not eliminate traditional forms of offending, it simply created an alternative environment in which they could occur.

These contrasting arguments also led to debate over the use of the terms cybercrime and computer crime. Though they were treated as synonymous during this period, they differed as to the function of technology in the course of the offense (Furnell, 2002; Wall, 2001). Computer crimes were thought to involve special use of computer technology in order to offend, while cybercrimes involved special knowledge of cyberspace (Furnell, 2002). This segmentation eventually disappeared in the mid-2010s, with the term cybercrime becoming dominant (Holt and Bossler, 2015). The use of cybercrime is particularly appropriate now, as virtually any and all devices are connected to the internet, including so-called Internet of Things devices like thermostats, home appliances, and even vehicles (Steinmetz and Yar, 2019).

Though cybercrime is now the dominant term, there are still some points of differentiation used by scholars, governments, and policy agencies. For example, academics

have critiqued the utility of the term cybercrime and presented alternatives to identify the unique nature of technological influences on behavior. McGuire (2007) argued for the use of the term hypercrime as a way to recognize that offenses occur as a result of the interconnected nature of virtual and real spaces which mirrors the World Wide Web itself. Others have argued in favor of terms like ecrime, technocrime, and informated crime as a means to avoid the unclear nature of what constitutes cyber in practice (see Steinmetz, 2023 for discussion). In fact, Cross (2019) argued we should abandon the term cyber when referring to all technology-facilitated fraud schemes, and simply note them as frauds as it otherwise limits the allocation of resources to properly combat these offenses.

The terms cyber-dependent and cyber-enabled crimes are used by some policymakers in the UK, Europe, and Australia as a means to identify how the offense occurs (Holt and Bossler, 2015; McGuire and Dowling, 2013). Cyber-dependent crimes are those that require a computer or technology in order to occur, such as the distribution of malicious software that affects the operations of computer hardware and software (McGuire and Dowling, 2013). Cyber-enabled crimes are those that are simplified by the use of computers and the internet, such as fraud and identity-based crimes (McGuire and Dowling, 2013). The use of cyber-dependent and cyber-enabled crimes have not been popularized or used consistently across place, making their overall utility limited.

Forms of cybercrime

While cybercrime in an incredibly important term in criminology, it should be noted that it is an umbrella term encompassing a wide range of offenses. There are various typologies of cybercrime proposed by researchers, though their acceptance is generally inconsistent (see Holt and Bossler, 2015 for review). One of the only typologies that gained acceptance in the field was created by David Wall (2001), which recognized four categories of offenses: 1) cyber-trespass; 2) cyber-deception and theft; 3) cyber-porn and obscenity; and 4) cyber-violence.

Acts of cyber-trespass are those that involve crossing boundaries of ownership in online spaces. For instance, a password-protected WiFi hotspot is effectively establishing a boundary of access, where only those with the password are allowed to use the connection (Holt and Bossler, 2015). Individuals who attempt to access these networks by guessing the password or using tools to break that password without authorization from the owner are effectively engaging in acts of trespass. Cyber-trespass activities are most often associated with computer hackers, as they often attempt to gain access to networks, email accounts, and other protected systems without permission (Jordan and Taylor, 1998; Steinmetz, 2016). Though not all hackers engage in such illegal activities, those who perform such acts cause significant financial harm to individuals, as well as public and private entities around the world (Holt and Bossler, 2015).

The second category involves acts of cyber-deception and theft which is complex and overlaps with acts of cyber-trespass (Wall, 2001). The term deception is used as criminals can readily misrepresent their identity while online through the use of fake names, email accounts, and stolen legitimate user credentials to access sensitive (Button and Cross, 2017; Leukfeldt et al., 2017). For example, hackers can utilize various methods to gain access to protected networks in order to obtain sensitive personally identifiable information and financial data (Hutchings and Holt, 2015; Leukfeldt et al., 2017; Roks et al., 2021). Actors then sell this information to others in online black markets in order to engage in fraudulent financial transactions, such as purchasing goods and services without the account holder's permission (Hutchings and Holt, 2015).

This category also includes independent acts of fraud that are enabled by email, social media, text messaging, and other forms of computer-mediated communication. For instance, there has been substantive criminological focus on the problem of romance scams, where offenders create false profiles on social media and dating applications and message potential targets in the hopes of generating a response (Button and Cross, 2017; Whitty, 2013). Once in contact, the offender then tries to develop an emotional and romantic rapport with the victim by asking personal questions and noting their love and attraction to them. As the relationship intensifies, the offender often asks the victim for financial support to travel to see them, or claim they have experienced a financial hardship that they cannot work through (Button and Cross, 2017). Such financial requests continue until such time as the victim ceases to support their requests. Victims of romance frauds frequently report intense emotional and psychological consequences as a result of their experiences with fraudsters (Button and Cross, 2017).

The act of illegally copying, sharing and/or obtaining digital media files, including computer software, audio, video, and e-books, without the permission of the copyright holder is also included in this category (Gopal et al., 2004). Digital piracy can be performed in a variety of ways, such as file sharing services, and is extremely common globally (Brown, 2016; Brown and Holt, 2018 for review). In fact, evidence suggests that most young people have engaged in piracy and easily justify their activities on the basis of the lack of harm caused to copyright holders and the costs associated with the legitimate purchase of media (Brown and Holt, 2018; Ingram and Hindjua, 2008).

The third category in Wall's (2001) typology of cybercrime involves cyber-porn and obscenity, reflecting a wide variety of sexual expression in online spaces. Sexually explicit content featuring adults, commonly defined as pornography, is readily available online and has become a major industry (Lane, 2000; Quinn and Forsyth, 2013). Such content can be produced with great ease due to the availability of high definition digital cameras, high-speed internet connectivity, and editing software. The consumption of adult sexual content is generally defined as legal in most nations, though it may be considered socially unacceptable depending on local mores (Quinn and Forsyth, 2013).

Technology has also transformed the nature of sex work in general (see Sex Work by Rand). The rise of social media and subscription-based media channels like Patreon

and OnlyFans has also allowed individuals to produce sexual content for a fee (Moloney, 2019; Rana, 2020). Similarly, sex workers utilize the internet to advertise sexual services that take place in offline settings, which may be illegal depending on the laws within a given country (Campbell et al., 2019). Their customers frequently use web forums and various online platforms as a tool to discuss their experiences, identify specific providers, and warn others about police and other efforts to curb sex work (Sanders et al., 2018).

The ease with which individuals can produce sexualized images and content has also led to individuals sending images of themselves to prospective or current romantic partners via direct text message or email. This practice has become colloquially known as sexting, and is popular as it is perceived as a somewhat secure method for sharing content to attract or stimulate a partner (Henry and Powell, 2018). Sexual content is not, however, secure, as the recipient can now readily share sexted images and video with others via text or social media, or on websites designed to embarrass or harass the sender (Henry and Powell, 2018). Such activities constitute revenge porn, and may be performed after a relationship ends, or by individuals who hack individual accounts to obtain access to their content (Henry and Powell, 2018).

This category also includes the sharing of child sexual abuse materials (CSAM), or images, video, and other media featuring images of people under the age of 18 engaging in sexualized acts (Eke and Seto, 2023). The distribution of CSAM is a global problem, and has been criminalized in virtually all nations. This content is shared on virtually all platforms online, which makes it exceedingly difficult for police to disrupt (Seto et al., 2018). In addition, some use the internet, particularly social media, as a means to connect with children in order to groom them into offline relationships, or to generate CSAM themselves (Eke and Seto, 2023).

The final category of Wall's (2001) typology of cybercrime encompasses acts of cyber-violence, recognizing the potential of the internet to allow individuals to engage in harmful behavior to others through online spaces. The most common behaviors included involve the use of social media, email, text, and other forms of online communication to bully, harass, threaten, or stalk others online (Vogels, 2021). The severity and frequency of messages an individual receives can lead victims to experience emotional harm, including embarrassment, shame, and may even lead some to physical harm through suicidal ideation (Wilson et al., 2022).

In addition, terrorists and extremist groups also use the internet as a means to recruit others into their movement, and promote their beliefs publicly and privately (Herath and Whittaker, 2021). The expression of hate speech and violence toward minoritized groups in society has become commonplace, particularly on social media (see Hate Crime and Networked Hate by Powell, Stratton, and Cameron). Actors also utilize various technologies as a means to coordinate and plan acts of offline violence (Hamm and Spaaij, 2017).

Furthermore, some extremist groups use the internet as a means to engage in cyberattacks against individuals, governments, and businesses. There is evidence that actors associated with animal and earth liberation groups have performed various cyber-

attacks over the last decade (Holt et al., 2021). Similarly, the hacker group Anonymous used DDoS attacks in order to stage protests against government, industry, and civilian targets (Olson, 2013). This has led some scholars to use the term hactivism to refer to cyberattack activities that may be deemed an act of protest or dissent by the actors (see Holt and Bossler, 2015) Thus, the use of technology has expanded the capability of extremist groups to affect populations and targets well beyond their overall capacity in offline environments.

Conclusion

Taken as a whole, cybercrime as a term reflects a wide range of offenses which differ in the degree to which they are facilitated by technology and impact people or computers and data. These differences highlight the inherent weaknesses of the use of the phrase, as virtually any act of crime or delinquency with some connection to computers and the internet could be classified as cybercrime. Though other ways of defining such offenses have been proposed, few have been able to supplant the socially accepted nature of the phrase (Steinmetz, 2023). It is possible that the concept of 'cyber' will become outmoded with continued technological advancements, which may enable the abandonment of the term in favor of the more realist approach of treating these offenses simply as crimes (Cross, 2019).

Suggested reading

Holt, T. J., & Bossler, A. M. (2015). *Cybercrime in Progress: Theory and Prevention of Technology-Enabled Offenses.* London: Routledge.

Holt, T. J., & Lavorgna, A. (eds.) (2021). *Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches.* Springer Nature.

Lusthaus, J. (2018). *Industry of Anonymity: Inside the Business of Cybercrime.* Cambridge, MA: Harvard University Press.

Steinmetz, K. F., & Yar, M. (2019). *Cybercrime and Society.* London: Sage.

Wall, D. (2007). *Cybercrime: The Transformation of Crime in the Information Age* (Vol. 4). Cambridge: Polity.

References

Brown, S. C. (2016). Where do beliefs about music piracy come from and how are they shared? An ethnographic study. *International Journal of Cyber Criminology*, 10(1), 21 – 39.

Brown, S. C., & Holt, T. J. (eds.) (2018). *Digital Piracy.* London: Routledge.

Button, M., & Cross, C. (2017). *Cyber Frauds, Scams and Their Victims.* London: Routledge.

Campbell, R., Sanders, T., Scoular, J., Pitcher, J., & Cunningham, S. (2019). Risking safety and rights: online sex work, crimes and 'blended safety repertoires'. *The British Journal of Sociology*, 70(4), 1539 – 1560.

Cross, C. (2019). Is online fraud just fraud? Examining the efficacy of the digital divide. *Journal of Criminological Research, Policy and Practice*, 5(2), 120 – 131.

Eke, A. W., & Seto, M. C. (2023). Correspondence of child age and gender distribution in child sexual exploitation material and other child content with age and gender of child sexual assault victims. *Sexual Abuse*, 35(3), 375 – 397.

Furnell, S. (2002). *Cybercrime: Vandalizing the Information Society.* London: Addison-Wesley.

Gopal, R. D., Sanders, G. L., Bhattacharjee, S., Agrawal, M., & Wagner, S. C. (2004). A behavioral model of digital music piracy. *Journal of Organizational Computing and Electronic Commerce*, 14(2), 89 – 105.

Grabosky, P. N. (2001). Virtual criminality: Old wine in new bottles? *Social & Legal Studies*, 10(2), 243 – 249.

Hamm, M. S., & Spaaij, R. (2017). *The Age of Lone Wolf Terrorism.* New York: Columbia University Press.

Herath, C., & Whittaker, J. (2021). Online radicalisation: Moving beyond a simple dichotomy. *Terrorism and Political Violence*, 35(5), 1 – 22.

Henry, N., & Powell, A. (2018). Technology-facilitated sexual violence: A literature review of empirical research. *Trauma, Violence, & Abuse*, 19(2), 195 – 208.

Hollinger, R. C., & Lanza-Kaduce, L. (1988). The process of criminalization: The case of computer crime laws. *Criminology*, 26(1), 101 – 126.

Holt, T. J., & Bossler, A. M. (2015). *Cybercrime in Progress: Theory and Prevention of Technology-Enabled Offenses.* London: Routledge.

Holt, T. J., Stonhouse, M., Freilich, J., & Chermak, S. M. (2021). Examining ideologically motivated cyberattacks performed by far-left groups. *Terrorism and Political Violence*, 33(3), 527 – 548.

Hutchings, A., & Holt, T. J. (2015). A crime script analysis of the online stolen data market. *British Journal of Criminology,* 55(3), 596 – 614.

Ingram, J. R., & Hinduja, S. (2008). Neutralizing music piracy: An empirical examination. *Deviant Behavior,* 29(4), 334 – 366.

Internet World Stats. (2023). Internet usage statistics. https://www.internetworldstats.com/stats.htm

Jordan, T., & Taylor, P. (1998). A sociology of hackers. *The Sociological Review*, 46(4), 757 – 780.

Lane, F. S. (2001). *Obscene Profits: Entrepreneurs of Pornography in the Cyber Age.* London: Routledge.

Leukfeldt, R., & Holt, T. J. (eds.) (2019). *The Human Factor of Cybercrime.* London: Routledge.

Leukfeldt, E. R., Kleemans, E. R., & Stol, W. P. (2017). Cybercriminal networks, social ties and online forums: Social ties versus digital ties within phishing and malware networks. *The British Journal of Criminology*, 57(3), 704 – 722.

McGuire, M. (2007). *Hypercrime: The New Geometry of Harm.* Abingdon and New York: Routledge-Cavendish.

McGuire, M., & Dowling, S. (2013). *Cyber Crime: A Review of the Evidence. Summary of Key Findings and Implications.* Home Office Research Report, 75, 1 – 35.

Maimon, D., & Louderback, E. R. (2019). Cyber-dependent crimes: An interdisciplinary review. *Annual Review of Criminology*, 2, 191 – 216.

Mann, D., & Sutton, M. (1998). NETCRIME: More change in the organization of thieving. *The British Journal of Criminology*, 38(2), 201 – 229.

Moloney, A. (2019). What are premium Snapchat accounts and are they just porn? *Metro*, February 21, 2019. https://metro.co.uk/2017/11/21/what-are-premium-snapchat-accounts-7088201/

Olson, P. (2013). *We Are Anonymous.* London: Random House.

Payne, B. K., and Hadzhidimova, L. (2020). Disciplinary and interdisciplinary trends in cybercrime research: An examination. *International Journal of Cyber Criminology*, 14(1), 81 – 105.

Perrin, A., & Anderson, M. (2019). *Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018.* Pew Charitable Trust. [Online] Available at https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

Quinn, J. F., & Forsyth, C. J. (2013). Red light districts on blue screens: A typology for understanding the evolution of deviant communities on the internet. *Deviant Behavior*, 34(7), 579 – 585.

Rana, P. (2020). Top 10 celebrity OnlyFans accounts: Cardi B and Bella Thorne to Tyga, here's all the steamy footage you need. *Meaww*, October 20, 2020. https://meaww.com/cardi-b-bella-thorne-tyga-shea-coulee-top-10-accounts-adult-site-only-fans-subscribe-what-they-post

Roks, R. A., Leukfeldt, E. R., & Densley, J. A. (2021). The hybridization of street offending in the Netherlands. *The British Journal of Criminology*, 61(4), 926 – 945.

Sanders, T., Scoular, J., Campbell, R., Pitcher, J., & Cunningham, S. (2018). *Internet Sex Work: Beyond the Gaze.* Cham: Springer International Publishing.

Seto, M. C., Buckman, C., Dwyer, R. G., & Quayle, E. (2018). *Production and Active Trading of Child Sexual Exploitation Images Depicting Identified Victims.* National Center for Missing and Exploited Children. https://calio.dspacedirect.org/handle/11212/4669

Steinmetz, K. F. (2016). *Hacked: A Radical Approach to Hacker Culture and Crime* (Vol. 2). New York: NYU Press.

Steinmetz, K. F., (2023). *Against Cybercrime: Toward a Realist Criminology of Computer Crime.* London: Routledge.

Steinmetz, K. F., & Yar, M. (2019). *Cybercrime and Society.* London: Sage.

Vogels, E. A. (2021). *The State of Online Harassment.* Pew Research Center. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/01/PI_2021.01.13_Online-Harassment_FINAL-1.pdf

Wall, D. S. (1998). Catching cybercriminals: Policing the internet. *International Review of Law, Computers & Technology*, 12(2), 201 – 218.

Wall, D. S. (2001). Cybercrimes and the internet. In D. S. Wall (ed.), *Crime and the Internet* (pp. 1 – 17). London: Routledge.

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(4), 665 – 684.

Wilson, C., Sheridan, L., & Garratt-Reed, D. (2022). Examining cyberstalking perpetration and victimization: A scoping review. *Trauma, Violence, & Abuse*, 24(3), 2019 – 2033.

Zickuhr, K. (2011). *Generations Online in 2010.* Pew Internet and American Life Project. Available at: www.pewinternet.org/Reports/2010/Generations-2010/Overview.aspx

Meropi Tzanetakis

18 Darknet

Abstract: This chapter explores the dual nature of the darknet, providing insights for students and criminologists alike. It examines how the darknet serves as a digital environment for both protecting privacy and freedom of expression and facilitating criminal activities, all rooted in a shared desire to evade commercial and state surveillance. By analyzing the enabling and constraining elements of the Janus-faced darknet, the chapter illustrates these dynamics through a detailed discussion of darknet drug distribution platforms, drawing on the author's original research.

Keywords: platforms, user data, encryption, privacy, surveillance, illicit drugs

The term darknet or dark web initially suggests something mystical, criminal, and threatening. In fact, however, darknet is not referring to the legal status of the content, but rather how certain services can be accessed on the internet (Tzanetakis, 2018). By crawling and analyzing darknet pages, Avarikioti et al. (2018) found that approximately 60 % of the content is legal, with illegal services including fraud, counterfeit offers, and drug markets. More broadly, darknet can be distinguished from the surface web or clearnet, which is accessible by search engines, and the deep web, which requires the user to take additional steps such as logging in or making a payment to access the content.

The internet can be imagined as an iceberg in the ocean with its visible tip being the surface web but only accounting for roughly about 10 % of the content. The much larger part of the iceberg is below the ocean and is referred to as the deep web with its content being invisible as it is inaccessible for search engines. The deep web comprises intranets of companies or organizations, as well as user accounts, among other elements. The darknet is then a small subset of the deep web that contains hidden services that are only accessible with specialized software.

To get access to the hidden services, encryption software like the TOR Browser (The Onion Router), Freenet, I2P (Invisible Internet Project) is required. In regular internet traffic, data moves between a user and its destination directly, often leaving identifiable traces. The darknet, however, facilitates network anonymity by routing the internet traffic through a series of randomly chosen servers. This process conceals the user's IP address and location, complicating law enforcement interventions. Technically, the darknet consists of a variety of darknets. In the following, the term darknet is used collectively for all darknets. By making use of TOR, users can communicate almost anonymously, ensuring no data is disclosed that could reveal their identity or whereabouts. However, the term darknet is misleading, as it implies that all content may be legally restricted, which is not necessarily the case as it contains a variety of content from file sharing to pornographic material, and from leaked data to drug markets.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-019

From privacy and freedom to surveillance and control

The darknet presents a Janus-faced digital environment, intriguing beyond its inherent mystery. It encompasses both privacy and freedom while entangled with elements of control and surveillance. On one side, it offers heightened privacy and anonymity, fostering uncensored communication and information exchange. Its encryption offers shelter for activists, journalists, and dissidents seeking to evade censorship, surveillance, and oppressive regimes. Conversely, it functions as a platform facilitating illicit activities, enabling illicit trade and criminal exchanges under the veil of anonymity. Initially, this duality presents a paradox: while providing refuge for the marginalized, it also facilitates persons engaged in criminal activities, encapsulating the complexity of the darknet. This supposed duality makes the darknet a compelling and intricate subject for criminological exploration. In this chapter, I will disentangle the darknet by arguing that both privacy and freedom of expression—regardless of whether it involves supporting fundamental human rights or criminal activities—have the same origin: they aim to circumvent and resist commercial or state control and surveillance. I will illustrate the Janus-faced darknet with my own research on darknet drug distributing platforms (Tzanetakis and South, 2023).

When browsing through the surface web or deep web, internet users leave digital traces, encompassing identifiable elements like IP addresses, cookies, and cache data, vital for website personalisation (Tzanetakis and Marx, 2023). Additionally, browser history, search queries, and login information contribute to a digital footprint. These traces form a digital trail, enabling websites to personalize content and facilitate tailored user experiences. However, they also raise significant privacy concerns as the traces can be utilized for targeted advertising, potentially invasive tracking, or even surveillance, emphasizing the impact of online activities on user privacy. They underscore the intricate balance between personalization and the potential exploitation of user data for commercial or monitoring purposes.

User-generated content is subject to monitoring, collection, and analysis by various actors, including the digital information economy. The design of digital platforms encourages their users to engage in prosumption practices which means that individuals function as both producers and consumers of content within a platform (Srnicek, 2017). These practices include activities such as uploading and sharing of content with others and using a 'like' button, tweet, or hashtag while swiping through text, video, audio, and pictures produced by other prosumers. Digital platforms engage in prosumption practices with the aim of monetizing the production and consumption of information, seeking to derive financial value from the digital data generated by users (Zuboff, 2019). With the advent of new practices and technologies, surveillance has become widespread and pervasive.

In contrast to commercial surveillance which focuses on collecting consumer data for marketing, profit, or service personalization, state-driven surveillance is primarily

conducted by government or law enforcement agencies, aiming to monitor and control citizens for reasons such as national security, law enforcement, and social control (Bauman and Lyon, 2013). The extent of surveillance on a global scale was uncovered by the Snowden files which revealed, among other things, mass data collection from private phone records and internet activities of both citizens and leaders worldwide (Lyon, 2014). The revelations demonstrated a critical breach of individual privacy and civil liberties, showing how citizens, irrespective of suspicion, were subjected to mass surveillance. This undermined the fundamental principles of democracy, fostering an environment of suspicion and eroding trust between governments and the general public. The broad and non-targeted data collection raised ethical concerns, highlighting the potential for abuse and the imbalance between security measures and personal freedoms.

The convergence of state-driven and commercial digital surveillance, aimed at behavioral regulation, control imposition, and profit generation, serves as a fertile environment for the emergence of the darknet. Within this digital realm, the darknet thrives by offering heightened privacy, freedom of expression, and access to information, enabling users to navigate digital spaces and communicate via digital media with increased anonymity. This, however, supports both fundamental human rights and criminal activities by enabling individuals to avoid censorship and disguising users' digital traces. Both share a common genesis by aiming to bypass and oppose control and surveillance by commercial industry or governmental bodies. Fundamental human rights and criminal activities initially appear to be diametrically opposed. In the following, I will use the example of my own original research on darknet drug platforms to disentangle this paradox.

Researching darknet drug platforms

Darknet drug platforms or cryptomarkets are technological infrastructures mediating exchange of a wide range of illicit drugs, among other products and services, between vendors and consumers (Tzanetakis and Marx, 2023). These digital platforms have been described as a "transformative criminal innovation" (Aldridge and Décary-Hétu, 2014) combining anonymizing technologies (e. g., TOR browser) with cryptocurrencies (e. g., Bitcoin) as a decentralized and non-government-issued means of payment (Barratt and Aldridge, 2016). In combination, these technologies enable its users to separate their identity from their cryptomarket activity, e. g., browsing through the offers, creating an account, or ordering illicit drugs.

The affordances of the darknet have contributed to a "transparency paradox" (Tzanetakis et al., 2016). While the platform operator organizes the marketplace, hosts multiple sellers, and defines the basic terms and conditions, sellers and buyers remain anonymous or pseudonymous while publicly organizing market exchanges. Acknowledged as unprecedented, the quality and availability of this data have introduced novel sets of data available to researchers (Barratt and Aldridge, 2016).

The darknet is a dynamic and ever-evolving environment on the internet (Tzanetakis and South, 2023). Characterized by frequently changing websites and a fluid landscape, it operates with an inherent volatility. While the content on the darknet spans the whole continuum of legality and illegality, particularly when associated with clandestine activities, websites in this realm are often short-lived and may involve sophisticated scams. The swift adaptations and constant shifts in online spaces pose challenges for research (as well as law enforcement), requiring a flexible approach to navigate the rapidly changing environments of illicit online activities (see Researching Online Forums by Šupa). Drug markets per se are also a fragmented and fluid research field, which changes over time and in reflection of societal developments, in order to avoid attention of the police.

Digital ethnography involves immersing oneself in online environments, to understand social phenomena from the participants' perspectives (Hine, 2015; Tzanetakis, 2021). While the method of digital ethnography is flexible and adaptable, it underscores the importance of contextualizing data in our globalized, translocal, and digitally interconnected world, considering both local meanings and the impacts of internet commercialization (see Online Ethnography by Gibbs and Hall).

The process involves an initial phase of inhabiting the digital environment to explore the hard-to-reach communities of darknet drug platform users (Tzanetakis, 2019). These communities consist of platform operators, customers, dealing customers, small-level dealers, high-level dealers, moderators, information hubs, law enforcement, researchers, journalists, and interested people who are browsing through platforms and discussion forums. For my research, this meant engaging in participant observation, including the user-generated content by self-presentations of vendors as well as monitoring of interactions between customers, operators, moderators, and vendors on various cryptomarkets and associated discussion forums. I also took screenshots and wrote field notes about my observations and perceptions. For the researcher, immersion serves the purpose of learning how to understand local meanings and familiarizing themselves with the specific cultural behaviors and norms.

The next digital ethnographic phase involves getting meaningful access to the hard-to-reach online community (Kaufmann and Tzanetakis, 2020). Field entry is influenced by following the cultural norms and rituals of the specific online community (see Accessing Online Communities by Kaufmann). For marginalized people to open up, a researcher needs to position themself as a cultural insider or outsider or both at the same time. Attaining cultural insider status necessitates the researcher sharing commonalities, roles, or experiences that resonate with the community under exploration. Adler and Adler (1987) identified three membership roles for qualitative researchers: peripheral, active, and complete. While Barratt and Maddox (2016) performed an active membership role in the darknet drug community with Barratt volunteering as administrator of a drug harm-reduction forum that was respected within the community, I adopted a peripheral membership role enabling me to simultaneously inhabit the roles of an insider and an outsider. A critical consideration for researchers lies in the choice between adopting a pseudonym, often advantageous for obtaining institu-

tional ethical approval, and entering the field openly, potentially promoting access and trust. However, opting for a pseudonym raises ethical dilemmas regarding transparency and authenticity within the research process. Balancing these factors is essential for maintaining research integrity and navigating the complexities of building relationships within the community studied.

My lack of complete immersion in the social world studied resulted in my being perceived as a cultural outsider. Simultaneously, I positioned myself as a cultural insider by acquiring the technological knowledge standards prevalent in the hard-to-reach darknet community. As I gained knowledge about the digital technologies employed, the operational dynamics of the ecosystem of darknet platforms, insights into vulnerabilities concerning law enforcement and fraud, as well as familiarity with the specific language and abbreviations, I consistently upheld a transparent researcher position. Moreover, forum moderators played an important role as gatekeepers to reassure the hard-to-reach online communities that my research was legitimate. Attaining complete membership within the darknet drug market community raises ethical concerns in research and may, in certain instances, cast doubt on the scientific validity of the research results. This role may involve the researcher actively engaging in unlawful activities. However, it provides the most comprehensive access to the community under study.

As part of this digital ethnographic phase, I was able to establish a credible online presence with the research participants in order to gain meaningful access to the darknet drugs community (Kaufmann and Tzanetakis, 2020). For my research, data collection included qualitative interviews with members of the hard-to-reach darknet drug market community, screenshots of interactions between community members on drug platforms and related discussion forums, screenshots of vendor shops, customer ratings, and field notes on observations of social practices.

Ethical considerations

Empirical research on darknet drug platforms comes with a variety of ethical considerations and challenges such as if the data available can be considered public or private, gaining participant's informed consent, ensuring anonymity, and protecting both researchers and research participants. And although this chapter cannot cover all ethical implications related to cryptomarket research, other chapters in this handbook as well Martin and Christin (2016) and Lavorgna and Holt (2021) are well positioned to further explore the multifaceted ethical challenges. Nevertheless, I would like to briefly address some key research ethical dimensions here. The affordances of the darknet present distinct challenges that deviate slightly from conventional ethical standards.

My research on darknet drug market practices includes sensitive research questions about activities that are considered criminal offenses in most jurisdictions. Accordingly, de-anonymizing participants can have legal consequences, even if this happens accidently. At the same time, establishing trust with community members is

essential when conducting digital ethnography. Here, the knowledge I gained about encryption software and specific settings that prevent detection by law enforcement are pertinent. Using suitable 'operational security' measures including non-traceable software (Jabber/XMPP messaging service with Off-the-Record messaging encryption) helped me to establish credibility in the field. In my interviews, I made sure not to collect identifying information (e.g., names, locations) and anonymize the data as soon as possible afterwards. This demonstrates that ethical considerations require navigating a delicate balancing to ensure the protection of research participants.

Thus, researchers first need to develop a sufficiently nuanced understanding of the specific digital environment, adhere to legal and regulatory policies including ethical guidelines of professional organizations such as the Association of Internet Researchers (AoIR) (franzke et al., 2020). Secondly, ethical decisions should be context-dependent, meaning they vary depending on the social and cultural context in which information is shared and exchanges occur (Nissenbaum, 2009). Thirdly, Markham (2018) introduced an impact model of ethics which emphasizes the potential consequences and effects of one's research. It involves evaluating how anything we do in the name of research is directly relevant to the community studied. Rather than merely harvesting digital interactions for our research aims, the focus should be on designing studies that enact social change within communities. This approach extends to considering how research influences the shape of possible futures, highlighting a holistic and socially responsible perspective on the impact of our research endeavours.

Disentangling the paradox of the darknet

At the first sight, darknet drug markets could be perceived as a threat due to their role in facilitating the anonymous buying and selling of illicit drugs, challenges in law enforcement efforts, and challenges for traditional drug control strategies. However, a more holistic perspective suggests that the growth of cryptomarkets is associated with a potential to reduce a variety of drug related harms (Aldridge et al., 2018; Bancroft, 2017; Barratt et al., 2016; Tzanetakis and von Laufenberg, 2016). Harm reduction is a public health approach that aims to minimize negative consequences, such as overdose and disease transmission, without necessarily advocating for abstinence (Lenton and Single, 1998). It includes strategies like needle exchange programs, supervised consumption rooms, and education to promote safer practices.

Darknet drug markets facilitate harm reduction by offering transparency on drug quality and purity of the drugs compared to traditional distribution channels. This factor holds significance because certain harms are linked to the potential for adulteration and the broader composition of drugs, leading to undesired effects or the risk of overdose. Another harm reduction benefit is minimizing physical and psychological violence for users compared to traditional transactions. Facilitated through features like the escrow payment system, the platform operator manages funds between transaction parties, preventing fraud and fostering trust. The operator serving as an inter-

mediary, this practice ensures financial oversight until the drugs are delivered as agreed upon. A third harm reduction aspect revolves around peer-to-peer information exchange on discussion forums and via rating systems to promote safer practices. They are used among peers to share experiences and information about the qualities of the drugs, effects, dosage recommendations, and polydrug use. Yet, these harm reduction aspects also involve surveillance activities entangled with encrypted data traffic. While harm reduction efforts aim to enhance safety, the encrypted data traffic also allows for the tracking and analysis of user interactions, transactions, and information exchange. This surveillance could potentially be carried out by various actors, including law enforcement agencies or other private actors interested in monitoring activities on these platforms.

A nuanced understanding of cryptomarkets reconciles the perceived contradiction of the darknet—a space for both criminal activities and marginalized free expression. Distributing drugs on darknet platforms is criminal, yet this environment minimizes harm from punitive measures. Amid the dual nature of the darknet, the emergence of new criminal offenses and expanding surveillance warrant careful consideration, balancing privacy, civil liberties, and the complex dynamics of this Janus-faced digital realm.

Suggested reading

Gehl, R. (2018). *Weaving the Dark Web: Legitimacy on Freenet, Tor, and I2P.* Cambridge, MA: MIT Press.
Lavorgna, A. (2020). *Cybercrimes: Critical Issues in a Global Context.* London: Macmillan.
Martin, J., Cunliffe, J., & Munksgaard, R. (2019). *Cryptomarkets. A Research Compendium.* Bingley: Emerald
Mirea, M., Wang, V., & Jung, J. (2019). The not so dark side of the darknet: A qualitative study. *Security Journal*, 32, 102 – 118.

References

Adler, P., & Adler, P. (1987). *Membership Roles in Field Research.* Newbury Park, CA: Sage.
Aldridge, J., & Décary-Hétu, D. (2014). Not an 'Ebay for drugs': The cryptomarket 'Silk Road' as a paradigm shifting criminal innovation. DOI: 10.2139/ssrn.2436643
Aldridge, J., Stevens, A., & Barratt, M. J. (2018). Will growth in cryptomarket drug buying increase the harms of illicit drugs? *Addiction*, 113(5), 789 – 796.
Avarikioti, G., Brunner, R., Kiayias, A., Wattenhofer, R., & Zindros, D. (2018). Structure and content of the visible Darknet. arXiv preprint arXiv:1811.01348.
Barratt, M. J., & Aldridge, J. (2016). Everything you always wanted to know about drug cryptomarkets* (*but were afraid to ask). *International Journal of Drug Policy*, 35, 1 – 6.
Bancroft, A. (2017). Responsible use to responsible harm: Illicit drug use and peer harm reduction in a darknet cryptomarket. *Health, Risk & Society*, 19(7 – 8), 336 – 350.
Barratt, M. J., & Maddox, A. (2016). Active engagement with stigmatised communities through digital ethnography. *Qualitative Research*, 16(6), 701 – 719.

Barratt, M. J., Ferris, J., & Winstock, A. R. (2016). Safer scoring? Cryptomarkets, social supply and drug market violence. *International Journal of Drug Policy*, 35, 24 – 31.

Bauman, Z., & Lyon, D. (2013). *Liquid Surveillance: A Conversation.* Cambridge: Polity Press.

franzke, a.s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* https://aoir.org/reports/ethics3.pdf.

Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday.* London: Bloomsbury.

Kaufmann, M., & Tzanetakis, M. (2020). Doing internet research with hard-to-reach communities: Methodological reflections on gaining meaningful access. *Qualitative Research*, 20(6), 927 – 944.

Lavorgna, A., & Holt, T. J. (eds.) (2021) *Researching Cybercrimes: Methodologies, Ethics and Critical Approaches.* Cham: Palgrave Macmillan.

Lenton, S., & Single, E. (1998). The definition of harm reduction. *Drug and Alcohol Review*, 17(2), 213 – 220.

Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 1 – 12.

Markham, A. N. (2018). Afterword: Ethics as Impact—Moving from error-avoidance and concept-driven models to a future-oriented approach. *Social Media + Society*, 4(3).

Martin, J., & Christin, N. (2016). Ethics in cryptomarket research. *International Journal of Drug Policy*, 35, 84 – 91.

Mayorga-Gallo, S., & Hordge-Freeman, E. (2017). Between marginality and privilege: Gaining access and navigating the field in multiethnic settings. *Qualitative Research*, 17(4), 377 – 394.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford: Stanford University Press.

Srnicek, N. (2017). *Platform Capitalism.* Cambridge: Polity Press.

Tzanetakis, M. (2018). The darknet's anonymity dilemma. In A. von Humboldt Institute for Internet and Society (ed.), *Encore 2017. The Annual Magazine on Internet and Society Research* (pp. 118 – 125). Berlin: HIIG.

Tzanetakis, M. (2019). Informal governance on cryptomarkets for illicit drugs. In A. Polese, A. Russo, & F. Strazzari (eds.), *Governance Beyond the Law: The Immoral, The Illegal, The Criminal* (pp. 343 – 361). Basingstoke: Palgrave Macmillan.

Tzanetakis, M. (2021) Qualitative Online-Forschungsmethoden: Digitale Datenerhebung und ethische Herausforderungen. In B. Prainsack & M. Pot, *Qualitative und interpretative Methoden in der Politikwissenschaft* (pp. 130 – 141). Wien: Facultas.

Tzanetakis, M., Kamphausen, G., Werse, B., & von Laufenberg, R. (2016). The transparency paradox. Building trust, resolving disputes and optimising logistics on conventional and online drugs markets. *International Journal of Drug Policy*, 35, 58 – 68.

Tzanetakis, M., & Marx, S. (2023). The dark side of cryptomarkets: Towards a new dialectic of self-exploitation within platform capitalism. In M. Tzanetakis & N. South (eds.), *Digital Transformations of Illicit Drug Markets: Reconfiguration and Continuity* (pp. 141 – 154). Bingley: Emerald.

Tzanetakis, M., & South, N. (eds.) (2023). *Digital Transformations of Illicit Drug Markets: Reconfiguration and Continuity.* Bingley: Emerald.

Tzanetakis, M., & von Laufenberg, R. (2016). Harm Reduction durch anonyme Drogenmärkte und Diskussionsforen im Internet? In akzept e.V. Bundesverband für akzeptierende Drogenarbeit und humane Drogenpolitik: 3. Alternativer Drogen- und Suchtbericht 2016 (pp. 189 – 194). Lengerich: Pabst Science.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* London: Profile Books.

Joanna Redden

19 Data justice

Abstract: This chapter provides an introductory overview of how data justice is being used as a key concept, framework, and research practice. The overview provides a condensed review of previous research that both informs and makes use of data justice as a research lens. The chapter concludes by highlighting how data justice approaches are particularly relevant for criminology.

Keywords: data, datafication, social justice, AI, automated decision systems

At the Data Justice Lab we use the term *data justice* as a key concept, framework, and practice to research the social justice implications of datafication (see Datafication by Chan), with a particular focus on changing power dynamics. While I am writing this entry, this overview is a condensed version of how we explain our use of the term in the co-authored introduction to our book *Data Justice* (Dencik et al., 2022). Data justice begins from the position that the increasing and widespread collection and use of data across different domains of social and political life is transforming ways of knowing and acting, decision-making processes, critical infrastructure and services, as well as working practices. The datafication of life is happening in ways that entrench and accelerate unjust and oppressive systems and discriminatory practices, while also introducing new power dynamics. For example, automated hiring systems have been found to discriminate on the basis of gender, mental health, disability, and ethnicity (Whitaker et al., 2019, Sánchez-Monedero et al., 2020). Facial recognition systems have been found to discriminate by misidentifying people with darker skin tones and women at rates much higher than white men (Bulamwini and Gebru, 2018). These systems have already led to wrongful arrest and detention. Automated decision-making systems (ADS) are leading to harm as people are falsely flagged as high risk through the use of automated scoring systems in child welfare, policing, sentencing, hiring, credit rating, and housing decision-making (Eubanks, 2018; O'Neil 2016; Redden et al., 2020).

While there are widespread discussions about datafication that promote its contributions to efficiency, security, or economic growth, too often such discussions fail to acknowledge the ways that increasing uses of AI and automated systems are already leading to harm. To prioritize data justice is to recognize that the expansion of data collection and use can and is entrenching and amplifying pre-existing injustices while introducing new power dynamics requiring investigation, critique, and resistance. By using data justice as a framework, the work we have been doing together at the Data Justice Lab stresses the importance of social justice concerns about datafication. Data justice in our work to date has been used as a lens to engage with and advance social justice as well as to challenge injustices related to datafication. Our use of

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-020

the term is informed by our collaborations as well as the significant and growing body of work by others.

Overview

Data justice has been informed by work highlighting how already marginalized groups are disproportionately negatively affected by emerging data systems, in ways that amplify and accelerate pre-existing inequality, discrimination, and oppression (Gangadharan et al., 2014; Noble, 2018; Benjamin, 2019; Hoffmann, 2019). It is also informed by media and digital justice work. The term digital justice has been used to link contemporary work to the longer history of media and digital reform (see Digital by Wernimont). As argued by Shade (2023), an early framework for digital justice was created by the Detroit Digital Justice Coalition which brought together community members, researchers, educators, artists, and entrepreneurs to advocate for equitable access to media resources, shared civic participation and decision-making about technologies, common ownership, co-operative models of development, and stresses the need to prioritize healthy communities. Taylor's data justice framework provides a means to account for oppression and different social contexts while highlighting the importance of autonomy and the ability to opt out of data collection (2017). Richard Heeks (2017) provides a framework to assess power and profit asymmetries in datafied applications. Sasha Costanza-Chock's (2018) *Design Justice* work centers on design processes and the conditions and contexts influencing the ways that data infrastructure is developed. This work calls for more participatory design practices that involve communities and building alternative bottom-up infrastructures to empower marginalized communities. For example, DiscoTechs (Discovering Technology Community Fairs) where participants learn things like media making skills, communication tactics, and how to build technology (Costanza-Chock, 2018).

Conceptions of data justice are informed by mobilization and resistance to datafied injustice. Given the lack of transparency and accountability surrounding much corporate and government uses of AI and automated decision-making systems (ADS), the work by grassroots groups, civil society organizations, and journalists to investigate, challenge, and advocate for more just societies and practices influences how we both understand and operationalize data justice. As just a few examples, the Center for Media Justice in the United States has created a Data Justice Lab in relation to Data for Black Lives to bridge research, data, and movement work. The Detroit Digital Justice Coalition have developed guidelines for equitable data practices. The American/Canadian Environmental Data Governance Initiative has developed an environmental justice framework. The Indigenous Data Sovereignty movement, made up of the alliance of groups all over the world, asserts Indigenous ownership about the collection and use of data by and about Indigenous peoples and lands. This movement builds on long-standing struggles against the extraction and exploitation of Indigenous peoples, their knowledge systems, customs, and lands (Kukutai and Taylor, 2016; Lewis,

2020). Data Justice is also being shaped by the growing platform co-operatism movement, which is challenging the monopolistic and extractive nature of platform capitalism.

The empirical research investigations we have done together at the Data Justice Lab informs our use of the term and has involved bringing together concerns about social justice, tech justice, and political mobilization (Dencik et al., 2016) and researching how datafication is changing state citizen relations and governance (Hintz et al., 2019; Redden, 2018). This work has also involved researching where and how people are already being negatively affected by data practices as well as learning from the people who are trying to redress and prevent data harms (Redden et al., 2017, 2020). Foregrounding the politics of data and attending to contexts has involved recognizing the importance of reparation and challenging cognitive injustices that fail to recognize non-mainstream and non-Western ways of knowing (Milan and Treré, 2019; Treré, 2019).

Prioritizing social justice in our research is based on recognition that the way technology is developed and used is not inevitable or neutral, but that technologies are embedded within social actors and forces and a particular political economy. Further, we share the view that technologies should be viewed as 'sites of struggle,' and that with any application there will be competing values, politics, and visions informing people's ideas about if and how technologies should be deployed (Eubanks, 2018). Attending to where such struggles are taking place reinforces the important fact that our datafied present and future are not predetermined, but that we have important work to do together to determine the kinds of datafied worlds we want to live in. Prioritizing social justice in this work involves recognizing our shared interdependence with each other and our environments.

Approaches to data justice are informed by long-standing concerns about the social justice implications of the nature of information and communication technologies. This work stresses the importance of understanding the way in which systems of measurement and evaluation, like the map and the census, provide a means through which people and territory have been surveyed and regulated (Kitchin, 2014; Hacking, 2007; Isin and Ruppert, 2019). There is a long history of states consolidating power through their uses of data and technology (Milner and Traub, 2021). Data then, about the world, has real-world impact as people use it to justify or make decisions. Information power, as argued by Hoffmann, is central to other forms of power (2020). For example, categorizations of people as deserving or undeserving has long been used as a way to justify or deny people's access to services or opportunities (see Categorization and Sorting by Franko). With AI and automation such processes of categorization and value are being embedded in infrastructure design and code. Data justice approaches are often informed by previous work outlining how data systems are part of longer histories and embedded in contexts where there is already a lot of structural and systemic violence and inequality. AI and ADS systems can intensify and accelerate this inequality because they incorporate classification and bureaucratic regimes that have been developed in ways that provide security to some but lead to greater vulnerability for

many (Spade, 2011). Key work in surveillance studies has detailed the technological shifts that enable increasing corporate and government surveillance and the need to be alert to how this contributes to social sorting and normalizing cultures of surveillance (Lyon, 2015, 2002; Gandy, 1993). This work stresses the importance of attending to datafication beyond concerns about how it enables privacy infringements, to concerns about the violations of human rights, new modes of subjectification, accelerated discrimination, pre-emption of due process in decision-making, and shifts in the ways governments come to know about and engage with people and social issues.

Data justice and criminology

Data justice is particularly relevant to criminology because many new AI and ADS applications are being used across criminal justice systems, an area where there are already high levels of inequality. For example, previous research documents inequality, systemic and structural racism in criminal justice systems in the United States, Canada, Australia, New Zealand, the United Kingdom, and Europe (see list below). This makes the use of technologies that could intensify and accelerate inequality and discrimination dangerous. Investigations of AI and ADS use in policing and child welfare have detailed how some of the most highly weighted variables in the systems analyzed, such as postal codes or whether or not a person receives benefits, are proxies for ethnicity or poverty and can lead to a disproportionate targeting and impact on already marginalized communities (Eubanks, 2018; O'Neil, 2016). These issues are leading to increasing challenges to uses of predictive systems for policing. For example, in April 2020 the Los Angeles Police Department announced they would stop using PredPol, this after having also stopped using a system called Laser (Redden et al., 2022). While the LAPD said they stopped the use of both in connection with COVID-19 financial constraints, representatives from the Stop LAPD Spying Coalition that had been investigating these systems, their impact, as well as challenging police use of the systems view the announcements as a response to years of community mobilization and organization. The internal audit by the Office of the Inspector General pointed to a number of issues including lack of oversight, inconsistency, and racial profiling. A civilian oversight panel questioned the effectiveness of the LAPD's systems. Stop LAPD's investigations raised significant concerns about the targeting of already marginalized people, particularly the unhoused (Redden et al., 2020).

Predictive technologies are also being used to risk assess the likelihood of a defendant re-offending and these scores are informing decision-making about sentencing and bail (see Sentencing and Risk Assessments by Ugwudike). In the United States, investigators working for ProPublica compared the risk scores assigned to 7,000 people and checked to see how many went on to be charged with a new crime. They found that the risk scores were unreliable and biased. The system was much more likely to falsely flag Black defendants as future criminals than white defendants. The system was also wrongly labeling white defendants as low risk more often than Black defendants (Ang-

win et al., 2016). The bias and errors produced in these kinds of systems are connected to the systemic biases embedded in the data being drawn upon to develop the systems (Lum and Isaac, 2016). Further, there is a growing body of work detailing how AI and ADS systems may violate people's rights, in ways that affect social rights, rights to due process and a fair trial, to the protection of personal data, and to privacy (see Privacy and Data Protection by Bygrave). Our research into cancelled systems demonstrates that systems are being legally challenged in multiple countries on different legal bases (Redden et al., 2022). For example, automated systems have been successfully challenged in different countries for different reasons including failed automation, not ensuring individual determination, a lack of due process, violating Europe's General Data Protection Regulation, being non-compliant with the European Convention of Human Rights and in contravention of the constitutional right to privacy (Redden et al., 2022: 18). This work also identified cities across the United States making the decision to pre-emptively ban uses of facial recognition technologies because of concerns about rights, discriminatory impact, and harm. These efforts are often connected to critical media coverage and civil society and community mobilization.

Conclusion

In our work at the Data Justice Lab we approach data justice as a research agenda that enables us to bring together the key social justice issues that come from our work. However, what data justice means is continually being shaped and developed. In this sense we do not understand data justice as an idea with a predefined end goal, or as an abstract ideal that can have universal applications or fixed practices associated with it (Dencik et al., 2022). In our Data Justice work at the Lab, we plan to continue to use it as a means of employing systematic analysis to understand the social justice implications of the increasing uses of data, AI, and automated decision-making systems. We see our approach with data justice as an avenue through which we can ask questions about how society is organized and the role technology has within it, so that we can also ask questions about how things could be otherwise to enable flourishing for everyone.

Suggested reading

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code.* Cambridge: Polity Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, PMLR, 81, 77 – 91.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need.* Cambridge, MA: MIT Press.

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data Justice.* London: Sage.

Eubanks, V. (2018). *Automating Inequality.* New York: Macmillan.

Milner, Y., & Traub, A. (2021). *Data Capitalism + Algorithmic Racism. Data for Black Lives and Demos.* Retrieved from www.demos.org/sites/default/files/2021 – 05/Demos_% 20D4BL_Data_Capitalism_Algorithmic_Racism.pdf

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Crown Publishing.

Richardson, R. (2021). Racial segregation and the data-driven society: How our failure to reckon with root causes perpetuates separate and unequal realities (2021). *Berkeley Technology Law Journal*, 36(3), 1051 – 1090.

Research documenting inequality, systemic and structural racism in AI and automated decision-making systems applications

US

https://www.sentencingproject.org/publications/un-report-on-racial-disparities/; https://www.ohchr.org/EN/NewsEvents/Pages/systemic_racism.aspx

CA

https://www.canada.ca/en/department-justice/news/2021/06/addressingsystemic-racism-against-indigenous-peoples-in-the-justice-system-government-of-canada-investments-andinitiatives.html

AU

https://www.alrc.gov.au/publication/pathways-to-justice-inquiry-into-the-incarcerationrate-of-aboriginal-and-torres-strait-islander-peoples-alrc-report-133/; https://www.sbs.com.au/nitv/nitv-news/article/2018/03/14/institutional-racism-evident-australian-justice-system

UK

https://warwick.ac.uk/services/sg/si/diversity/advance_he_-_understanding_racism_report.pdf

EU

https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690525/EPRS_BRI(2021)690525_EN.pdf

NZ

https://www.hrc.co.nz/news/migrant-newzealanders-experience-fear-loss-culture-and-identity-due-racism/

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica.* Retrieved from www.propublica.org/article/machine-bias-risk-assessmentsincriminal-sentencing.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code.* Cambridge: Polity Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, PMLR, 81, 77 – 91.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need.* Cambridge, MA: MIT Press.

Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of antisurveillance resistance in political activism. *Big Data & Society*, 3(2), 1 – 12. doi: 10.1177/2053951716679678

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data Justice.* London: Sage.

Eubanks, V. (2018). *Automating Inequality.* New York: Macmillan.

Gandy, O. (1993). *The Panoptic Sort: A Political Economy of Personal Information.* New York: Harper Collins.

Gangadharan, S. P., Eubanks, V., & Barocas, S. (2014). *Data and Discrimination: Collected Essays.* Open Technology.

Hacking, I. (2007). Kinds of people: Moving targets. *Proceedings of the British Academy*, 151, 285 – 318.

Heeks, R. (2017). A structural model and manifesto for data justice for international development. *Development Informatics Working Paper Series, No. 69.*

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2019). *Digital Citizenship in a Datafied Society.* Cambridge: Polity Press.

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of anti-discrimination discourse. *Information, Communication & Society*, 22(7), 900 – 915.

Hoffmann, A. L. (2020). Terms of inclusion: Data, discourse, violence. *New Media & Society*, 23(12): 3539 – 3556.

Isin, E., & Ruppert, E. (2019). Data's empire: Postcolonial data politics. In D. Bigo, E. Isin, & E. Ruppert (eds.), *Data Politics: Worlds, Subjects, Rights* (pp. 207 – 227). Abingdon: Routledge.

Kitchin, K. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.* London: Sage.

Kukutai, T., & Taylor, J. (eds.) (2016). *Indigenous Data Sovereignty: Toward an Agenda.* Canberra: ANU Press.

Lewis, J. E. (ed.) (2020). *Indigenous Protocol and Artificial Intelligence Position Paper.* Honolulu, Hawai'i: The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR).

Lum, K., & Isaac, W. (2016). To predict and serve. *Significance*, 13(5): 14 – 19. Retrieved from https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

Lyon, D. (2002). *Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination.* New York: Routledge.

Lyon, D. (2015). *Surveillance after Snowden.* Cambridge: Polity Press.

Milan, S., & Treré, E. (2019). Big data from the South (s): Beyond data universalism. *Television & New Media*, 20(4): 319 – 335.

Milner, Y., & Traub, A. (2021). *Data Capitalism + Algorithmic Racism. Data for Black Lives and Demos.* Retrieved from https://www.demos.org/sites/default/files/2021-05/Demos_%20D4BL_Data_Capitalism_Algorithmic_Racism.pdf

Noble, S. U. (2018). *Algorithms of Oppression.* New York: New York University Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Crown Publishing.

Redden, J., Brand, J., & Terzieva, V. (2017, 2020) *Data Harm Record.* Cardiff University: Data Justice Lab. https://datajusticelab.org/dataharm-record/

Redden, J., Brand, J., Sander, I., & Warne, H. (2022). *Automating Public Services: Learning from Cancelled Systems.* Carnegie UK Trust. https://carnegieuktrust.org.uk/publications/automating-public-services-learning-from-cancelled-systems/

Redden, J. (2018) Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data and Society*, July, 1 – 13. https://doi.org/10.1177/2053951718809145

Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 Conference on Fairness, Aaccountability, and Transparency* (pp. 458 – 468).

Shade, L. (2023). From media reform to data justice. In M. Gallagher & A.V. Montiel (eds.), *The Handbook of Gender, Communication, and Women's Human Rights* (pp. 71 – 87). Wiley Online Books. https://doi.org/10.1002/9781119800729.ch5

Spade, D. (2011). *Normal Life: Administrative Violence, Critical Trans Politics and the Limits of the Law.* Boston, MA: South End Press.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), 1 – 14. doi: 10.1177/2053951717736335

Treré, E. (2019). *Hybrid Media Activism: Ecologies, Imaginaries, Algorithms.* Abingdon: Routledge.

Whitaker, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., Morris, M. R., Rankin, J., Rogers, E., Salas, M., & West, S. M. (2019). *Disability, Bias, and AI.* AI Now. https://ainowinstitute.org/wp-content/uploads/2023/04/disabilitybiasai-2019.pdf

Rocco Bellanova

20 Databases

Abstract: Databases underpin much of our digital societies, and participate in how we are governed. Yet, it remains challenging to grasp databases' socio-technical functioning and their role within practices of control. This chapter first resituates digital databases in relation to other techniques of governing. It then leverages on diverse disciplinary perspectives to identify promising research avenues for the study of digital databases, notably in the European domains of public security and border and migration controls.

Keywords: database interoperability, governing data, media theory, power–knowledge relations, practices of control

Introduction

A database, as suggested by anthropologist Tahani Nadim (2021: 125), is basically "a collection of related data organized to facilitate swift search and retrieval." Databases are not only about the storage of information, but also about supporting "knowledge discovery" and "deriving new information that did not exist before" (Revesz, 2010: 1). These initial definitions highlight one important aspect of databases' role in social control: a database translates a more or less complex reality into a pre-defined system of knowledge that some actors can then use to govern that and/or other social realities. In other words, the relation between a database and the realities we inhabit is one in which the former is very successful in shaping the latter, because it informs our worldview and how we attempt to control social realities.

It is difficult to think about a facet of our everyday, and of our digital societies at large, that is not influenced or shaped by the functioning of a digital database. Of course, other media operate, and have historically operated, in between reality and its control—think about how archives, repositories, account books, or museums have

Disclaimers and acknowledgements: Rocco Bellanova's research was funded by the European Union (European Research Council ERC Starting Grant DATAUNION, project no. 101043213). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Rocco would like to thank Mareile Kaufmann and Heidi Mork Lomell for their constructive comments, and Francis Hunger and Raphael Gellert for their valuable suggestions.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-021

been, and still are, used by institutions and corporations. Yet, the fact that most contemporary databases operate with digital data matters in terms of their impact, not least because "digitized data are transforming government knowledge practices" (Ruppert, 2012: 122). In addition, digital databases are essential to exercise what sociologist Michael Mann (1984: 192) calls "the infrastructural power of the state," because they become an increasingly important element of its "logistics of political control." It is therefore not surprising that, since the 1960s, digital databases have raised substantial worries about boosting state and corporate surveillance (Garfinkel, 2000). The adoption of the first data protection legislation in France in 1978, for example, can be seen as a legal and policy response to maintain control over databases (Poulain, 2022).

While many of us are aware of databases' omnipresence and importance, it is still challenging to completely grasp their socio-technical functioning and, ultimately, how they participate in governing us and our societies. This chapter offers some vantage points to explore the relation between databases and practices of control. It first re-situates the term conceptually in relation to a broader history of techniques of governing. It then unpacks how literatures in computer science and media theory define databases, so to foreground their complexity and socio-cultural effects. In a third section, the chapter discusses the European landscape of databases in the field of security and border and migration control, highlighting crucial dynamics at play there. The conclusions suggest some avenues for researching databases from a digital criminology perspective.

Databases between power and knowledge

Databases inform our lives. Whenever you book a flight, stream a movie, or cross a border, you are using a database or feeding one with information about you, your perceived behavior, or the transactions between you and the company, the device, or the border guards. At the same time, corporations and public authorities are using databases for business or security purposes to maximize their profits or to assess whether you are a security threat or are allowed to enter the country. Companies and institutions are often eager to build bigger databases, to get access to new ones, or to interconnect existing ones (for example, by making them interoperable and thus able to speak to each other). This is evident when thinking about social media platforms and leading IT companies like Microsoft, Google, or Meta (van Dijck et al., 2018). It can also be observed in those domains that are studied by criminology and other disciplines interested in the study of social control (Monahan and Palmer, 2009), such as public security or border and migration control.

Databases belong to a broader and longer history of governing through archival and quantification practices. For example, the history of statistics is intertwined with important transformations of statecraft across Europe at least since the late 17th century (Desrosières, 1998). Think also about the production and circulation of documents feeding the archives created by former colonial authorities, allowing European colonial powers to exert control at a spatial and temporal distance (Stoler, 2002).

Like statistics or archives, databases are crucial elements of what Michel Foucault (1995 [1977]: 27) calls "power-knowledge relations." These are the material and heuristic practices underpinning any "constitution of a field of knowledge," and any "power relations" (Foucault 1995 [1977]: 27). In other words, if knowledge is crucial to exercise power, cognitive and practical aspects of knowledge generation become themselves primary concerns for those who govern. Take, for example, the adoption of 'Bertillonage' by the French police in the late 19th century (Piazza, 2005). This was an identification system predating the large-scale use of biometrics as we know it. This policing tool did not only rely on photography and supposedly scientific knowledge about the human body, but also on the constitution of a complex information system and its management (Ellenbogen and Langmead, 2020). Or consider the account books used by "American and West Indian slaveholders from the late eighteenth century through the American Civil War" (Rosenthal, 2018: 2). These were very material technologies of knowledge and violence, which leveraged a managerial approach to further the exploitation of enslaved people, with "the soft power of quantification supplement[ing] the driving force of the whip" (Rosenthal, 2018: 2).

As media theorist and artist Francis Hunger (2018: 54) noted, when the term *database* emerged in the early 1960s, "it reflect[ed] the broader need to deal with an ever growing amount of information in industrial capitalism from 1900 onward." Hence, rather than insisting on their novelty, the question should be how digital databases and the diffusion of computers feed and support specific forms of power-knowledge relations or their transformation (see Computation by Mazzilli Daechsel and Datafication by Chan). For example, in their *Policing the Risk Society*, criminologists Richard Ericson and Kevin Haggerty (1997: 13) noted that "[c]omputers allow the development of new formats of risk communication, as well as instant dispersal of knowledge of risk to interested institutions," and that—in this context—it is "[t]he databases rather than the individual bureaucrat [that] become the basis for governance through knowledge." In another classic of criminological literature, David Garland highlighted the effects of IT systems and "computerized data processing" for the "systematization of criminal justice," and thus the ability of some public authorities to further control and direct other authorities (2001: 115). Researchers should then keep in mind that techniques akin to the digital database have been, and keep, underpinning practices of governing, control, and coercion (see, for example, on the role of lists in counter-terrorism: de Goede and Sullivan, 2016). It is thus important to ask what difference the digital nature of databases makes, when, and in relation to whom.

Databases: materiality, subjectivity and cultures

No database functions in isolation, but as part of a broader setting. As a computer science handbook puts it, a database is just one of the "four components" of a "database system," the other components being the "users, the database application, [and] the database management system (DBMS)" (Kroenke et al., 2020: 18). The database manage-

ment system and the database application are "computer programs," the former "used to create, process, and administer the database," and the latter "serv[ing] as an intermediary between the user and the DBMS" (Kroenke et al., 2020: 18). Computer sciences define the database itself as a "self-describing collection of related records" (Kroenke et al., 2020: 18). In practice, this means that a database should always include "a description of [its] structure" and therefore of its "contents" (Kroenke et al., 2020: 18). Such an understanding of the database emphasizes its actual complexity and highlights how both human and non-human actors are needed to make a database work. It also invites us to unpack the structural choices that define, for instance, what data a database system can collect, and how such data shall be stored in view of a pre-defined retrieval process. This material perspective can help us better grasp how technical decisions may, in fact, have far-reaching implications in terms of power-knowledge relations.

Digital databases are remarkable media, and it is not surprising that various strands of media theory have been engaging with them in the last three decades. Two approaches are worth mentioning here. First, Mark Poster's (1995: 78–94) understanding of "databases as discourse." That is, the power of the database is not only to operate as a mighty tool in the hands of companies or public authorities, but first and foremost to "effect a constitution of the subject" (Poster, 1995: 85). In other words, databases do not just represent a parcel of an objective reality out there, but they participate in transforming people into entities (subjects) that can be read and mobilized by those that govern. Compared to earlier technologies, a digital database facilitates social control by partially automating the "interpellation" of a given subject (Poster, 1995: 90). In insisting on the database's transformative power upon any person, this approach invites more classical forms of resistance, namely privacy and data protection, to be abandoned (see Privacy and Data Protection by Bygrave). From this perspective, a critique of the database grounded in human rights would fall short, either because it presupposes an "autonomous individual" that never was (Poster, 1995: 93), or because it misses the actual functioning of a database system as a machine geared towards the "extraction and production of data and meta-data," which is different from mere collection and storage (Hunger, 2018: 62).

A second perspective on databases is formulated by media theorist Lev Manovich (2002). Manovich (2002: 218 and 219) approaches the database as a "cultural expression," if not *the* "symbolic form of the computer age." Since the late 1990s, databases have become somewhat familiar objects, used for work or entertainment, well beyond the remit of database specialists. People have grown used to "perform[ing] various operations – view, navigate, search" (Manovich, 2002: 219). Still, a database requires other techniques to extract meaning—or actionable knowledge—from the information stored therein. By foregrounding the relation between user and database as a site in which cultures emerge, Manovich (2002: 226–228) invites us to consider the role of interfaces and algorithms. In other words, as databases become foundational to a given culture (including security or crime control cultures), we may want to study how various digital media (and not *just* the database or *just* the algorithm) relate to each other and affect power-knowledge relations. For example, we can ask about the implications

when the police start feeding data from pre-existing databases about criminal offenses into geographic information systems (GIS). Focusing on Chicago law enforcement, political geographer Brian Jordan Jefferson (2017) highlights the problematic use of GIS-generated visualizations of crime data. These novel data visualizations spearhead a cultural approach to crime control—the 'hotspot' one—that may in turn feed into the further marginalization of often already racialized populations (see Bias by Oswald and Paul).

European databases for policing, migration, and borders and their criminological relevance

In the European Union (EU), policing, migration, and border controls are practices that largely rely on the use of digital databases (see Borders and Border Control by Jeandesboz). The EU has set up several large-scale databases including millions of records concerning, among others, people asking for asylum, requesting a visa, or being suspected or flagged as criminals (Glouftsios, 2018). Other EU centralized databases are currently under construction, notably to control the entrance of travelers without an EU passport as well as the length of their stay. Along with national databases set up in EU member states, these European IT systems are expected to facilitate the exchange of data across national and institutional boundaries, and to supplement (and in some cases substitute) physical borders by setting up digital ones (Broeders, 2007). From the 1990s onwards, these supranational databases have been participating in "invent[ing] other ways to distribute, network and circulate authority," confirming at the same time the political saliency of crime, migration, and border controls even when it is not possible or desirable to create "a centralized policing superstate" (Walters and Haahr, 2005: 105).

European databases for policing, migration, and borders are massive systems. By the end of 2022, the oldest European database in this domain, the Schengen Information System (SIS II), contained more than 86 million alerts (eu-LISA 2023: 10)—that is, information about either persons or objects that authorities across Europe upload in a central database to facilitate surveillance across Schengen member states. Furthermore, several EU databases—including the SIS II itself—store biometric data. This is worth considering because biometrics are not only particularly sensitive information, but they also embody state authorities' eagerness to control a population via a supposedly unambiguous form of individual identification (Cole, 2002). In turn, this feeds into the ambition to control the mobility of people—first and foremost migrants—notably through the collection, storage, retrieval, and comparison of biometrics extracted from their bodies (Amelung et al., 2020: 15–34). As criminologist Katja Franko (2011: 336) puts it, this database-driven approach produces a pre-emptive differentiation between mobile people—some turned into "'crimmigrant' bodies" and others considered "bona fide travelers." This differentiation is largely irrespective of peoples' actual behavior

and further marginalizes those falling into the crimmigrant category (see also M'charek et al., 2013).

In 2019, the EU adopted legislation to make all centralized databases in the field of security, border, and migration interoperable. This ambitious project foresees the creation of complex technical systems to facilitate the networking of large-scale EU databases through a series of layers and interfaces, thus "formatting European security integration" (Bellanova and Glouftsios, 2022: 454), and shifting EU "governmental rationales [...] from identity production to identity management" (Leese, 2022: 115). This means that, besides the technical challenges, interoperability has important legal, political, and institutional effects (Bastos and Curtin, 2020). Interoperability is expected to foster and ease access by police, border, and other security authorities to data that were not initially collected for crime control, counterterrorism, or border security. Yet, from a legal perspective, this change puts into question data protection's capacity to maintain meaningful control over the large-scale processing of personal data, notably by countering the concentration of too much information into the hands of public authorities. Paraphrasing legal scholars Paul De Hert and Serge Gutwirth (2006: 28), the separation of databases partakes in maintaining the "separation of powers" and thus "keep[ing] those powers in check." Interoperability also transforms the institutional landscape of policing. Take, for example, the multiplication of so-called fusion centers in the United States, which organize and foster the exchange of data originally stored by diverse actors (public and private) in diverse repositories. Not only do these centers strengthen "the surveillance net in which people are caught," but they also reorient the relation among various public authorities, and between them and the private sector (Monahan and Palmer, 2009: 632).

Conclusions

Like other socio-technical practices such as archives (see Archives by Bonde-Thylstrup and Veel), account books, or statistics, databases are at the core of the nexus between knowledge and power. As computer science and media theory teach us, they are complex systems made of hardware, software, and human beings. At the same time, digital databases are also "core technologies that support and lie behind so many other systems" (Dourish, 2017: 105). Decidedly, they are at the core of much policing across Europe (and beyond), with database interoperability paving the way for the emergence of novel power–knowledge relations. Even though our current socio-political and scholarly attention seems to be mostly concerned with algorithms and Artificial Intelligence (AI) (see Artificial Intelligence by Van Brakel), unpacking the role of databases in our digital society remains paramount if we are to understand the inner workings of social control. For instance, algorithmic systems need databases to operate, and this is notably the case for many practices relevant to criminologists, such as profiling or intelligence-led policing. Even those AI systems that promise to operate without tra-

ditional databases, still need to collect and organize disparate datasets to generate new knowledge.

In the spirit of digital criminology, we can identify at least two other research lines that promise to remain relevant in the future, if not gain further socio-political and scholarly significance. A first line of research consists of approaching databases as lively and relational entities. In practice, this means engaging in the analysis of those maintenance practices (Glouftsios, 2021) that keep databases functioning in practice, or of the evolution of databases jointly governed by leading IT companies to detect harmful online content (GIFCT, 2024). A second promising field of research approaches digital databases not only as research objects but also as precious research tools. Some criminologists already rely on existing governmental databases to gather "alternative data sources for criminological research" (De Moor et al., 2017). Others engage in the creation of new databases, like the *Deaths at the Borders Database* (Last et al., 2017). This important initiative collects, in one repository, available but otherwise dispersed administrative data about migrants' deaths at the European borders, to make this potentially critical knowledge count in research and policy cycles.

Suggested reading

Agar, J. (2003). *The Government Machine.* Cambridge, MA: MIT Press.

Eubanks, V. (2018). *Automating Inequality.* New York, NY: Picador.

Vavoula, N. (2022). *Immigration and Privacy in the Law of the European Union – the Case of Information Systems.* Leiden: Brill Nijhoff.

References

Amelung, N., Granja, R., & Machado, H. (2020). *Modes of Bio-Bordering.* Singapore: Palgrave.

Bastos, F. B., & Curtin, D. (2020). Interoperable information sharing and the five novel frontiers of EU governance. *European Public Law*, 26(1), 59 – 70.

Bellanova, R., & Glouftsios, G. (2022). Formatting European security integration through database interoperability. *European Security*, 31(3), 454 – 474.

Broeders, D. (2007). The new digital borders of Europe. *International Sociology*, 22(1), 71 – 92.

Cole, S. A. (2002). *Suspect Identities: A History of Fingerprintings and Criminal Identification.* Cambridge, MA: Harvard University Press.

de Goede, M., & Sullivan, G. (2016). The politics of security lists. *Environment and Planning D*, 34(1), 67 – 88.

De Hert, P., & Gutwirth, S. (2006). Interoperability of police databases within the EU: An accountable political choice? *International Review of Law, Computers & Technology*, 20(1&2), 21 – 35.

De Moor, S., Vander Beken, T., & Van Daele, S. (2017). DNA databases as alternative data sources for criminological research. *European Journal on Criminal Policy and Research*, 23(2), 175 – 192.

Desrosières, A. 1998. *The Politics of Large Numbers.* Cambridge, MA: Harvard University Press.

Dourish, P. (2017). *The Stuff of Bits.* Cambridge, MA: MIT Press.

Ellenbogen, J., & Langmead, A. (2020). Forms of equivalence: Bertillonnage and the history of information management. *Technology and Culture*, 61(1), 207 – 238.

Ericson, R. V., & Haggerty, K. D. (1997). *Policing the Risk Society.* Oxford: Clarendon Press.
eu-LISA. (2023). *2022 SIS II Annual Statistics.* Tallinn: eu-LISA.
Foucault, M. (1995 [1977]). *Discipline & Punish.* New York, NY: Vintage Books.
Franko, K. (2011). 'Crimmigrant' bodies and bona fide travelers. *Theoretical Criminology*, 15(3), 331 – 346.
Garfinkel, S. (2000). *Database Nation.* Sebastopol, CA: O'Reilly.
Garland, D. (2001). *The Culture of Control.* Chicago: University of Chicago Press.
GIFCT. (2024). *GIFCT's Hash-Sharing Database.* Global Internet Forum to Counter Terrorism, accessed 08.12.2024. https://gifct.org/hsdb/.
Glouftsios, G. (2018). Governing circulation through technology within EU border security practice-networks. *Mobilities*, 13(2), 185 – 199.
Glouftsios, G. (2021). Governing border security infrastructures: Maintaining large-scale information systems. *Security Dialogue*, 52(5), 452 – 470.
Hunger, F. (2018). Epistemic harvest. The electronic database as discourse and means of data production. *A Peer-Reviewed Journal About*, 7(1), 52 – 65.
Jefferson, B. J. (2017). Digitize and punish: Computerized crime mapping and racialized carceral power in Chicago. *Environment and Planning D*, 35(5), 775 – 796.
Kroenke, D. M., Auer, D. J., Vandenberg, S. L., & Yoder, R. C. (2020). *Database Concepts. Ninth edition.* Hoboken: Pearson.
Last, T., Mirto, G., Ulusoy, O., Urquijo, I., Harte, J., Bami, N., Pérez Pérez, M., Macias Delgado, F., Tapella, S., Michalaki, A., Michalitsi, E., Latsoudi, E., Tselepi, N., Chatziprokopiou, M., & Spijkerboer, T. (2017). Deaths at the borders database: Evidence of deceased migrants' bodies found along the southern external borders of the European Union. *Journal of Ethnic and Migration Studies*, 43(5), 693 – 712.
Leese, M. (2022). Fixing state vision: Interoperability, biometrics, and identity management in the EU. *Geopolitics*, 27(1), 113 – 133.
M'charek, A., Schramm, K., & Skinner, D. (2013). Topologies of race. *Science, Technology, & Human Values*, 39(4), 468 – 487.
Mann, M. (1984). The autonomous power of the state. *European Journal of Sociology / Archives Européennes de Sociologie / Europäisches Archiv für Soziologie*, 25(2), 185 – 213.
Manovich, L. (2002). *The Language of New Media.* Cambridge, MA: The MIT Press.
Monahan, T., & Palmer, N. A. (2009). The emerging politics of DHS fusion centers. *Security Dialogue*, 40(6), 617 – 636.
Nadim, T. (2021). Database. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.), *Uncertain Archives* (pp. 125 – 132). Cambridge, MA: MIT Press.
Piazza, P. (2005). Alphonse Bertillon face à la dactyloscopie. Nouvelle technologie policière d'identification et trajectoire bureaucratique. *Les Cahiers de la sécurité intérieure*, 56, 251 – 270.
Poster, M. (1995). *The Second Media Age.* Cambridge: Polity Press.
Poulain, C. (2022). Le projet SAFARI (1970 – 1974). *Terminal*, 134 – 135, np.
Revesz, P. (2010). *Introduction to Databases: From Biological to Spatio-Temporal.* London: Springer.
Rosenthal, C. (2018). *Accounting for Slavery: Masters and Management.* Cambridge, MA: Harvard University Press.
Ruppert, E. (2012). The governmental topologies of database devices. *Theory, Culture & Society*, 29(4/5), 116 – 136.
Stoler, A. L. (2002). Colonial archives and the arts of governance. *Archival Science*, 2, 87 – 109.
van Dijck, J., Poell, T., and de Waal, M. (2018). *The Platform Society.* Oxford: Oxford University Press.
Walters, W., & Haahr, J. H. (2005). *Governing Europe.* Abingdon: Routledge.

Janet Chan

21 Datafication

Abstract: This chapter examines the origin and meaning of the term "datafication" and analyzes data practice in terms of the ideological and political nature of data, supportive and resistant narratives, and contexts and consequences of datafication.

Keywords: data, sociotechnical imaginaries, contextual factors, consequences

Datafication is a term coined by Mayer-Schönberger and Cukier in their book *Big Data* —"To datafy a phenomenon is to put it in a quantified format so that it can be tabulated and analyzed" (2013: 78). According to the authors, datafication should not be conflated with digitization (which merely converts analogue information into binary code so that it is in a computer-readable format). Jefferson (2020) suggests that police (at least in the United States) have started to represent events in the world in the form of data points (for example, in Geographical Information Systems) since the 1960s. Crime occurrence databases have existed even longer. In criminal justice, datafication has taken place not only in record keeping, but also in mapping, suspect identification, correlation, and prediction of crime occurrences (see Prediction by Ķīlis, Gundhus and Galis).

At the simplest semantic level, datafication is about turning something into "data." But what is "data"? Mayer-Schönberger and Cukier trace the word "data" to mean "given" in Latin, "in the sense of a 'fact'" but modern usage of the word "refers to a description of something that allows it to be recorded, analyzed, and reorganized" (2013: 78). In policing, Ratcliffe (2014: 99) has placed data along a DIKI (data, information, knowledge, intelligence) continuum, so that, for example, when the location of a burglary incident is recorded on a computer, it is regarded as *data*; when this incident is recognized (together with other data) by a crime analyst as an emerging pattern, it becomes *information*; when this information is shared with a detective and is used to build hypotheses about the "criminal environment," this new wisdom becomes *knowledge*; when this knowledge is used by police to mount a surveillance operation, this knowledge becomes *Intelligence* (see Intelligence by Gundhus and Lundgaard). Under this typology, datafication represents the first step in a cumulative process of transforming an incident (itself subject to processes of detection, definition, and decision to report) into a form that can be recorded, compared, and made sense of using human interpretation and/or analytical tools. Note that *Intelligence* is sometimes replaced by *Understanding* (and/or *Wisdom*) in different models (Quarmby and Young, 2010: 26–27).

With the advancement of analytic tools that can make sense of textual, sound, and visual signals, there is no reason to assume that datafication is only about *quantitative* data. For example, a frequency (count) of burglaries in a certain time period turns in-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-022

cidents into quantitative data, whereas photographs of stolen items—before they are labeled and counted—exist as qualitative data.

Datafication is more than a one-way process towards the accumulation of data. A productive way to conceptualize datafication is to regard it as a *practice* which involves two "interwoven processes: the use of *more and different data* … and the deployment of *more advanced methods* to analyze these data and feed it back into existing work processes" (Reutter, 2022: 905; original emphasis). To see datafication as practice has the advantage of appreciating it as both *dynamic* and *interactional:* datafication is constantly changing and its trajectory is dependent on both structural and symbolic conditions. Chan and Bennett Moses (2017), for example, make use of Pierre Bourdieu's framework to situate social practice as an interaction of the *field* (with its various types of *capital*) and the *habitus* (a generative mechanism based on shared assumptions about the nature and purpose of data) of security agents. For example, technological change can enhance the cultural capital of policing agents with technical expertise; it can also transform what agents regard as relevant information (Chan and Bennett Moses, 2017: 303). This framework implies that the practice of datafication in a particular field may be limited/facilitated by the presence/absence of structural constraints/resources as well as the nature of shared assumptions held by actors in the field.

Sheila Jasanoff's (2015) notion of sociotechnical imaginaries is useful for examining shared assumptions. Jasanoff posits that sociotechnical imaginaries are "collectively held, institutionally stabilized, and publicly performed visions of desirable futures" in relation to a specific technological change (2015: 4). This formulation acknowledges that "resistant imaginaries" can co-exist with "desirable" ones. By tracing how sociotechnical imaginaries are formed, resisted, negotiated, and implemented into "assemblages of materiality"—data assemblages that are not only manifested in values and thoughts, but also in "ownership structures, policy agendas, organizational practices, and legal frameworks"—in a particular administrative context (Reutter, 2022: 906), we are in a better position to identify the causes or mediators of harmful unintended consequences.

Thus, datafication practice can be understood in terms of (i) the ideological nature of data, (ii) the political nature of datafication, (iii) narratives that support datafication projects, (iv) narratives that resist datafication, (v) contextual factors that mediate or inhibit datafication, (vi) consequences of datafication, and (vii) data justice as a framework for action.

(i) *The ideological nature of data.* Data is more than the material presence of paper records or digital information stored on databases on physical or "cloud" storage. It is intrinsically ideological. Datafication relies on an ideology of *dataism* which assumes that data is objective and can be trusted for accurate tracking of human behaviors or social phenomena (van Dijck, 2014). In this sense, datafication is less about the production of data but more about the constitution of "symbolic and imaginative work" that makes it possible to "think of something as 'data' in the first place" (Dourish and Gómez Cruz, 2018: 2):

> Data do not speak for themselves. Data must be narrated—put to work in particular contexts, sunk into narratives that give them shape and meaning, and mobilized as part of broader processes of interpretation and meaning-making. (2018: 1)

Creating narratives is part of the process of *sensemaking* (Weick, 1995) that people engage in "to explicate the world and give it a sense of order" (Chan and Bennett Moses 2017: 302).

(ii) *The political nature of datafication.* Datafication is political in that it involves the privileging of certain ways of framing, selecting, and excluding information so that certain claims and discourses are legitimized or delegitimized (Jasanoff, 2017: 11–12; Chan et al., 2022). Dencik and Kaun (2020: 3) suggest that "Datafication … needs to be treated as a political development rather than as a technological one, and one that sets out methods of knowing and definitions of social life that have implications for what values, logics, and forms of responses are privileged over others."

(iii) *Narratives that support datafication.* Reutter (2022: 905) has found that datafication is central to visions about the future of the welfare state and public administration:

> Such sociotechnical imaginaries include notions of proactive, rather than reactive, modes of governance. Datafication provides the public sector with a sense of being able to do more, better, faster, and more cheaply and is therefore perceived as a solution to the growing complexity of society and administration and as a tool to reduce uncertainty …

Chan et al. (2022) have highlighted the symbolic attractions of data-driven policing approaches such as the "scientification" of police work, legitimizing discourse such as "smart" analytics and the likelihood of more "rational and objective" decision making.

(iv) *Narratives that resist datafication.* Critiques of datafication have arisen from many quarters. Some *police officers* see data-driven policing as "'deskilling' and 'devaluing' their experiential craft of policing … replace their discretion … [and] threaten the legitimacy of the case-based approach aimed a conviction" (Chan et al., 2022: 4). *Social scientists* have identified flaws such as "opaqueness, reinforcement of discrimination, and facilitation of surveillance" in datafication projects, while case studies have documented the unintended consequences of data-driven systems (Reutter, 2022: 904).

Chan (2021) points out that *data scientists* themselves can create narratives that resist datafication. In particular, D'Alessandro et al. (2017: 125–126) raise specific issues of discrimination that machine-learning systems can introduce, using predictive policing as a case study: sources of discrimination include data issues ("discrimination in, discrimination out"), systemic biases against certain social groups, sample bias (see Bias by Oswald and Paul), model misspecification, and process failure such as "inappropriate feedback loops" which could "lead to a self-perpetuating system that continuously targets poorer and more minority concentrated communities" (2017: 132).

Community groups who are affected by the adverse effects of datafication can also be authors of resistant imaginaries. Chan (2021: 55) has shown how community organizations, researchers, and other concerned parties in the US have expressed reservations about predictive policing, calling for a "more transparent, community-based and fair systems that are subject to informed public debate, independent evaluation and continuous monitoring as to their racial impact." In a highly controversial (and subsequently found illegal) application of data-driven fraud detection algorithm in Australia (the so-called Robotdebt Scheme which automated the income data verification and debt notification stages of welfare fraud detection), the resistant imaginary took the form of public criticisms and ultimately a class action lawsuit against the federal government. The class action resulted in a $112 million settlement in 2021 (see Commonwealth of Australia, 2023; Mann, 2020).

(v) *Contextual factors that mediate or inhibit datafication.* As Chan et al. (2022: 2) point out, the pace of datafication has been driven by "Society's demand for risk information …, technology's promise of organisational efficiency …, and the attractions of the rhetoric of 'intelligence-led policing' …". However, Reutter's (2022: 918) research shows "how policy, organizational structures, legal frameworks, subject matter experts, and existing data infrastructures are able to mediate datafication in significant ways. These constraints act as counterforces against dominant sociotechnical imaginaries strongly dominated by the private sector …." By the same token, the above-mentioned ill-conceived Robodebt Scheme in Australia was facilitated by an organizational culture that discourages the delivery of "bad news" to superiors, disregards questions about the scheme's legality, and presents obstacles to the giving of independent advice (Commonwealth of Australia, 2023: 124; see Privatization by Lomell). Research on the impact of datafication on the production of police intelligence has identified challenges such as the symbolic mismatch between police experiential knowledge and data-driven intelligence and deficits in material conditions such as technical knowledge and poor system integration (Chan et al., 2022: 11).

(vi) *Consequences of datafication.* Consequences of datafication may not become obvious until it has been implemented for some time, either through lack of independent monitoring/evaluation or difficulty of accessing reliable information. For example, the biases of the risk assessment tool COMPAS were not identified until Angwin et al. (2016), at least six years after it was rolled out by probation departments in New York State. Even though there was a "comprehensive statistical evaluation of the tool" published in 2012, racial differences were not evaluated. Similarly, the Robodebt Scheme in Australia started operating in 2015 and did not end until 2019. A robust assessment of its consequences was only published in 2023. A scheme that was designed to save $1.7 billion in its first five years had the opposite effect: some 380,000 individuals were affected and the government's subsequent reimbursement and writing off of debts amounted to $1.751 billion

(Commonwealth of Australia, 2023: 471; xxix). By the time the harms caused by ill-conceived schemes become known, it is usually too late to "fix" the problems.

(vii) *Data justice as a framework for action.* Data justice has emerged as an overarching framework for engaging with the consequences of datafication, even though the concept is itself "unsettled" (see Dencik and Sanchez-Monedero, 2022 for how different disciplines approach the analysis of and responses to data injustice). This framework recognizes that "the burdens of datafication overwhelmingly fall on resource-poor and marginalised groups in society" and invites us to "focus on what function datafication—as a discourse and practice—serves in different contexts, the social and political organisation that enables it, and who benefits" (2022: 9–10). Even though political actions may range from refusal, divestment of resources, strategic litigation, to a movement towards data sovereignty, the common ground is "a need to tackle the actual conditions that lead to experiences of injustice as they exist on the ground rather than necessarily pouring efforts into appealing to ideal formations of data and technology in contemporary society" (2022: 11). Such actions are to be "nurtured through solidarity" and social relations (2022: 11).

Conclusion

Viewing datafication as a form of *practice* that involves the translation of sociotechnical imaginaries into data assemblages provides opportunities for citizens, users, and researchers to "alter these imaginaries prior to or even during their translation" (Reutter, 2022: 918). Participating in the alteration or improvement of these imaginaries may take a number of paths. The search for data justice may be approached in different ways, depending on the circumstances and conditions that result in injustice.

- The contestation of dominant sociotechnical imaginaries may be an important step. It provides opportunities to find out what the obstacles are and how imaginaries can be changed.
- It is important to observe how sociotechnical imaginaries are negotiated by institutions, workers, and users and who are the winners and losers (Reutter, 2022).
- Participants must avoid "data-driven determinism"—instead, they should look at how "policy, organisational structures, legal frameworks, subject matter experts, and existing data infrastructures can mediate datafication in significant ways" (Reutter, 2022: 918).

More generally, researchers could investigate datafication practices in different settings using ethnographic methods such as "situational analysis" which identifies "all the discourses, objects and actors involved … as well as their relations and negotiations" and uncovers "how the structural, cultural and operational contexts … led to differing definitions and perceptions of technology" by different actors (see Sanders and Chan, 2023: 778–779).

Suggested reading

Chan, J., & Saunders, P. (eds.) (2021). *Big Data for Australian Social Policy: Developments, Benefits and Risks*, Academy of Social Sciences in Australia, Canberra, https://socialsciences.org.au/publications/big-data-for-australian-social-policy/

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London and New York: Routledge.

Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1428

McDaniel, J. L. M., & Pease, K. G. (eds.) (2021). *Predictive Policing and Artificial Intelligence.* London: Routledge.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias.* Pro-Publica, 23 May 2016 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Chan, J. (2021). The future of AI in policing: Exploring the sociotechnical imaginaries. In J. L. M. McDaniel and K. G. Pease (eds.) *Predictive Policing and Artificial Intelligence* (pp. 43 – 59). London: Routledge.

Chan, J., & Bennett Moses, L. (2017). Making sense of big data for security. *British Journal of Criminology*, 57, 299 – 319.

Chan, J., Sanders, C., Bennett Moses, L., & Blackmore, H. (2022). Datafication and the practice of intelligence production. *Big Data & Society*, 9(1), 1 – 13. DOI: 10.1177/20539517221089310

Commonwealth of Australia. (2023). *Report.* Royal Commission into the Robodebt Scheme.

d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious classification: A data scientist's guide to discrimination aware classification. *Big Data*, 5(2), 120 – 134. DOI: 10.1089/ big.2016.0048.

Dencik, L., & Kaun, A. (2020). Datafication and the welfare state. *Global Perspectives*, 1(1), 12912. https://doi.org/10.1525/gp.2020.12912.

Dencik, L., & Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, 11(1). https://doi.org/10.14763/2022.1.1615

Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, 5(2), 1 – 10. DOI: 10.1177/2053951718784083

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasanoff & S.-H. Kim (eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (pp. 1 – 33). Chicago: University of Chicago Press.

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightline of justice. *Big Data & Society*, 4(2), 1 – 15.

Jefferson, B. (2020). *Digitize and Punish: Racial Criminalization in the Digital Age.* Minneapolis: University of Minnesota Press

Mann, M. (2020). Technological politics of automated welfare surveillance: Social (and data) justice through critical qualitative inquiry. *Global Perspectives*, 1(1), 12991. https://doi.org/10.1525/gp.2020.12991.

Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data.* London: John Murray.

Quarmby, N., & Young, L. J. (2010). *Managing Intelligence: The Art of Influence.* Sydney: Federation Press.

Ratcliffe, J. H. (2014). *Intelligence-led Policing.* London: Routledge.

Reutter, L. (2022). Constraining context: Situating datafication in public administration. *New Media and Society*, 24(4), 903 – 921.

Sanders, C. B., & Chan, J. (2023). Methodological reflections on researching the sociotechnical imaginaries of AI in policing. In S. Lindgren (ed.) *Handbook of Critical Studies of Artificial Intelligence* (pp. 773 – 782). Cheltenham UK and Northampton, MA, USA: Edward Elgar Publishing.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197 – 208.

Weick, K. E. (1995). *Sensemaking in Organizations.* Thousand Oaks: Sage Publications.

Daniel Trottier

22 Digilantism

Abstract: This chapter considers the practice of digilantism, in which unaffiliated individuals seek online retaliation against seemingly illegal and immoral actions. Local communities and content creators adopt mobile devices and social platforms to make wrongdoers visible to an ever-expanding audience. These practices are inherently ambivalent. Not only do they implicate a range of progressive and regressive causes, they also complicate relations between states and vigilantes.

Keywords: digilantism, vigilante, justice seeking, scrutiny, denunciation

Introduction

Daniel Solove's 2007 book *Future of Reputation* opens with a now infamous 'dog poop' incident from South Korea, where a woman refused to pick up her dog's waste in a subway (see also Dennis, 2008 and Reichl, 2019). This triggered on-site and online outrage against her presumed sense of entitlement. Someone on the subway uploaded a photograph of her, at which point she was publicly identified. Popular culture dictates that the internet does not forgive, nor does it forget. This global media spectacle brought disproportionate punishment for the target, including online and offline harassment.

This case is one of the first major examples of digilantism, which refers to unaffiliated individuals taking offense at seemingly illegal and immoral actions and engaging in online retaliation. While digilantes may cooperate with police by sharing evidence, so too do they engage in their own forms of punishment, including getting the target fired, online harassment, and physical retaliation. Digilantism is justice seeking performed by individuals not deputized by local authorities. Participants use digital tools to identify and denounce alleged wrongdoers. These practices are criminologically relevant as a mediated expansion and mutation of embodied vigilantism (Johnston, 1996). Local communities and content creators adopt mobile devices and social platforms to make wrongdoers visible to an ever-expanding audience. In publishing allegations against a targeted individual, there is often a presumption of guilt that serves to bypass any deliberation or appeals to reason.

Funding: This research was funded by the Dutch Research Council (NWO), project number 276 – 45 – 004.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-023

Digilantism is ambivalent in that it includes both regressive and progressive mobilizations. It may address long overlooked harms like sexual abuse (see Abuse by McAlinden), but often exceeds what we might consider a proportionate response. Its mandate extends seamlessly from crimes to moral offences that are not (yet) considered criminal. Despite the digital moniker, it not only includes responses to online offenses like revenge porn and nuisances like 'voluntourism' (posting problematic representations of humanitarian work, see Schwarz and Richey, 2019), but also offensive offline behavior like animal abuse (Chang and Poon, 2017) and 'manspreading' (taking up too much space in public, see Jane, 2017). Moreover, offenses and annoyances can be denounced online as well as offline, as when posters identifying people engaging in hate speech are plastered all over a city centre (Plesničar and Šarf, 2020). Other terms used include digital vigilantism, cyber vigilantism, internet vigilantism, virtual vigilantism, netilantism, and the 'human flesh search engine' in the Chinese context.

Digilantism in relation to vigilantism

We can build upon Johnston's six dimensions of embodied vigilantism that include (i) planning and premeditation by (ii) private voluntary citizens that (iii) constitutes a kind of social movement of 'autonomous citizenship' that (iv) uses and threatens the use of force (v) against a (perceived/potential) social transgression, and therefore (vi) conveys a sense of security to a local public, despite being illegal in many jurisdictions (Johnston, 1996: 220). These characteristics are to some degree challenged or disrupted by digital media (Trottier, 2017). The most striking dimension that digitization challenges is arguably the use of force (iv), which now relies heavily on cultural violence in terms of weaponization of visibility. This is a lasting and searchable stigma, that other media actors can yield at any point in time. Planning and premeditation (i) is also complicated, as digilantism campaigns are held together by a diversity of social actors. Some participants may trigger a vigilante reaction when they impulsively photograph and share asocial behavior they witness on a train. Other participants may comment on and share this content as part of their routine social media practice (see Social Media by Twigt), without a particular movement (iii) or reassurance of security (vi) in mind. But their actions may be coached, coordinated, and enhanced by those who intentionally cultivate a digilante movement, or online following. These professional digilantes may cultivate a sense of mediated vigilance among their audiences (Trottier, 2020). They create a type of audience participation that not only reacts to footage that entrepreneurs put out on their channels, but also is willing and able to source offensive content for these channels. It is important to view digilantism not just as series of cases, but as a generalized set of practices on social media that combines consuming content and enacting some sense of justice and punishment of targets.

Employing a vigilantism framing allows scholars to make sense of new practices emerging out of digital media usage. This lens also helps underline that digilantism is often representative of hegemonic values within a society (Rosenbaum and Seder-

berg, 1974), but a means for marginal communities to cope with their own vulnerabilities (Madsen, 2004). Abrahams also recognizes that researchers and practitioners "have some choice in how narrowly or broadly we define" vigilantism (1998: 6). Criminologists should keep this ambivalence in mind when assessing digital versions. The fact that incidents are crowdsourced among social actors on a global scale means that what seems like a 'pure' case of digilantism on first pass will include other actors (e.g., tabloid media online) engaging in other practices (e.g., harassment, counter-denunciations).

Several typologies highlight the moral ambivalence of just and harmful cases of vigilantism. Loveluck (2020) proposes a classification based on four ideal types that include more peripheral practices like flagging, organized leaking, and core practices like investigating and hounding. He acknowledges that any single case may include several of these types, for instance when an attempt to identify an assailant may scale up to "a more sustained mobilisation against" the individual that includes harassment (Loveluck, 2020: 227). In response Galleguillos proposes "a conceptual distinction between attacks and investigations" (2022: 358). Even if these fold into each other easily in practice, it is useful to consider that these may be distinct phases that mobilize separate actors. Here, practices like citizen journalism or citizen science may overlap with digilantism when they generate content that is knowingly or unknowingly used to denounce a targeted person (Mano, 2010). Likewise, digilantes may frame their actions as citizen journalism as a means of legitimation. Gainford (2022) proposes four broader sets of practices, namely informative, investigative, punitive, and disruptive. These distinctions are useful for mapping diverse motivations. Once again, these are ideal types that are complicated, as even local events can bring in global actors to denounce and punish targets. Punitive digilantism in particular is "reliant on the audience's willingness to accept and consume the content" (Gainford, 2022: 203). This speaks to how digilantism depends on the good will of an audience to sustain legitimacy, as well as mediators and platforms that frame and share this content.

Relations with the state

Relations between vigilantes and the state are also important, especially when police may mobilize others to investigate. Relations between states, civilians, and private organizations are nodal (Shearing and Wood, 2003), as collaborations in policing are renegotiated. Researchers address relations between digilantes and police, noting that the former can assist in sharing information. Some collectives are "a potentially useful, if under-valued, component of cyber-security networks" (Huey et al., 2013: 81). High profile cases shape public perception, bringing demand for local authorities to be proactive when handling digilantism.

In the European context collaboration between police and public "are here to stay, but the rules of this 'game' have yet to be established," and the police "discursively set boundaries" between acceptable and unacceptable public engagement online (Dekker

and Meijer, 2020: 298). On the role of the state in Morocco, Chalfaouat (2020) notes digilantism is enacted through co-production of security in public participation. He recommends "the promotion of civic virtue to the public interest and the avoidance of retaliation" (Chalfaouat, 2020: 182). Digilantism is a continuum that includes collaborative relations with public authorities alongside practices deemed illegal. Researchers may consider how citizen (digital) vigilance is promoted by police and governments. Governments may mobilize civil society groups like StopXam, a digilante group against bad parking and other traffic nuisances (Gabdulhakov, 2020). Yet these partnerships may be abandoned and denounced by state media when no longer advantageous.

Digilantes are also motivated by the perception that police cannot handle digital crimes, such that they seek to restore order themselves. Likewise, research on a crowdsourced investigative Reddit community reveals "an ambivalent rapport between [community] members and the police whose limitations are frequently invoked to justify the development of investigative activities in the first place" (Myles et al., 2020: 333, *brackets added*). We can consider the diverse skills often required to perform digilantism, including hacking (see Hacking by Wall), open-source intelligence (OSINT) investigations, and the ability to command a supportive audience. Digilantes also rely on social engineering, including those who target so-called 'email scammers' (Byrne, 2013). These skills are centered around making the target visible while either remaining anonymous, or making oneself visible in a strategic way to maintain control over the media narrative. In these contexts, digilantes engage in image work when interacting with the public to "negotiate who they are (not) and what they (do not) do" (Myles et al., 2020: 319). Criminologists should follow how these practices are negotiated by digilantes, whether these are cohesive groups like neighborhood watch communities, or temporary partnerships on Reddit.

Digital media platforms also promote global tendencies in media practice, including in justice seeking. Most prominently, they become common venues for denouncing child sexual abuse in countries including the United States, Germany, the Netherlands, Cambodia, and Russia (Hadjimatheou, 2021). These global initiatives have local adaptations that reflect systemic harms. In the Russian context, so-called pedophile hunting 'safaris' willfully conflate pedophilia "with homosexuality under the category of 'nontraditional sexual relations'", enabling "moral entrepreneurs to use the almost unchallenged condemnation of child abuse to strengthen their broader moral and political goals" (Favarel-Garrigues, 2020: 306).

Progressive and regressive digilantism

Progressive forms of digilantism against, for example, sexual abuse emerge due to a perceived lack of legitimate and effective institutional alternatives, including being able to rely on the police. Disproportionate punishment by digilantes may be explained by public perception that these crimes are poorly handled by law enforcement. Public discourse that celebrates feminist digilantism risks framing these practices as the only

viable recourse for anyone wishing to avoid sexual abuse (Jane, 2016: 292). As digilantes are embedded in temporary partnerships with police, criminological research should address the division of responsibility between feminist digilantes and local authorities who may not want to appear unresponsive to gendered forms of abuse. Feminist digilantism is situated in post-#MeToo context: these initiatives raise awareness of long-lasting acceptance of sexual harms across countries and contexts. Yet, feminist digilantes frequently encounter regressive denunciations and abuse online. At the same time, digilantism that targets women more generally also prevails, including among migrant communities from Central Asia (Gabdulhakov, 2019). Women in China may also be denounced by being placed into established sexist tropes, including "the ungrateful traitor, the corrupt elite, and the ugly slut" (Huang, 2023: 1). Here, no tangible crime may be committed, but moral policing may blend into criminal policing. Users may take offense on nationalist and gendered grounds, at which point they search for additional reasons to denounce and punish targets.

Progressive reckonings also happen in response to police violence against minorities, which has historically failed to receive appropriate justice. This is more openly acknowledged in the wake of the Black Lives Matters movement. In addition to denouncing police, digilantes target individuals who engage in illegal or immoral actions against minorities (Williams, 2020). Yet anti-Black digilantism also pervades. Consider the anti-scamming digilantism of 419eater.com in which members reply to suspected Nigerian scammers by "invok[ing] more traditional antiblack vigilante aesthetics" (Byrne, 2013: 78). The same tools and practices are used to target both racist and anti-racist initiatives.

Digilantism often reflects political polarization in civilian mobilizations. Regressive and progressive movements invoke digilantism when making appeals to law and order, which can serve to promote broader social and political movements. It is important to note that for some digilante groups, "the core of their existence involves a physical presence in public space" (Vicenová, 2020: 249), such as those in Slovakia engaging in anti-Roma and anti-migrant vigilantism. The term 'societal vigilantism' suggests that for far-right participants, the "target is not deviant behaviours but collective attributes perceived as deviant" (Tanner and Campana, 2020: 263). Beyond selective 'crime fighting', far-right vigilantes position themselves as "soft leaders" within "information bubbles" (Tanner and Campana, 2020: 276). Selective 'crime fighting,' then, extends into social movements and collective identity construction (Johnston, 1996) to perpetrate violence against marginalized communities. Digilantism covers a range of contexts involving progressive and regressive extrajudicial attacks on wrongdoers. It can be used as tool to either minimize or scale up categorical harms. This leads to concerns about ethics over the appropriateness of digilantes doxing avowed fascists (Douglas, 2020; Milbrandt, 2020).

Another concern with the term digilantism is the relation and potential overlap with other digital practices. Digilantism coexists with digital activism: activists can easily shame and denounce targets as part of a broader repertoire. Targets may be individuals, but are likely to also include corporations and states. Conversely, digilantes may

label themselves as activists depending on their socio-political ambitions (Gabdulhakov, 2020). They may engage in activist practices, especially mediated and communicative ones. At the same time, they may continue to bring attention to perceived wrongdoers. This may be both for the sake of restoring order, and to legitimate their cause. Digilantism may also resemble cyber-bullying: denunciation of a target may be invoked as moral justification for bullying (cf. Loveluck, 2020). But cyber-bullies and other malicious actors may simply fabricate legal and moral charges against a target, for instance when engaged in 'swatting' (sending police to a target's house on false pretences). We can also consider that some participants in digilante campaigns may seek to abuse others under morally justifiable conditions.

Finally, we can ask what counts as participation in digilantism? This varies from initiating campaigns against targets and making physical interventions, to adding views, likes, and comments to a viral video, and thus joining a collective audience. Leading and contributing to digilante activities typically amounts to content generation practices (Favarel-Garrigues, 2020). We can extend the notion of participation to consider the role of media platforms and the press in directing mediated attention to cases of digilantism. Content creators, tabloids, and other media actors report and editorialize on the offense, the vigilante retaliation, and any social or legal fallout (Trottier, 2020). These incidents propel a digital media economy based on engagement. In turn, public understandings of crime and punishment are shaped not only by media entities, but in a way that narrows the focus to the alleged moral failings of individuals, and other often unnamed masses of individuals who are said to punish them.

Conclusion

The term digilantism evokes several forms of ambivalence that warrant further attention:

- Digilantism transcends an online/offline distinction, as incidents can be fully mediated on digital platforms, all while bearing material impact on local communities.
- While digilantism is often a response to criminal events, digital media users can also mobilize against moral offences with the same repertoires.
- These pursuits may reflect progressive as well as regressive notions of justice, notably in polarized contexts.
- Police and other state actors may choose to support or to sanction these endeavors.
- Taken together, digilantism remains a troubling mix of justice seeking and entertainment.

Crime and entertainment media have a long history of being intertwined, through various initiatives that include crime-based reality television. Scrutiny and denunciation of others spills across contexts, including fandoms. Not only does denunciation become a form of entertainment, but entertainment itself becomes denunciation entertainment when anti-fans pursue justice via cancellation of celebrities (Ng, 2022).

Suggested reading

Loveluck, B. (2020). The many shades of digital vigilantism. A typology of online self-justice. *Global Crime*, 21(3 – 4), 213 – 241.

Nhan, J., Huey, L., & Broll, R. (2017). Digilantism: An analysis of crowdsourcing and the Boston marathon bombings. *British Journal of Criminology*, 57(2), 341 – 361.

Trottier, D. (2017). Digital vigilantism as weaponisation of visibility. *Philosophy & Technology*, 30, 55 – 72.

Trottier, D., Gabdulhakov, R., & Huang, Q. (eds.) (2020). *Introducing Vigilant Audiences.* Cambridge: Open Book Publishers.

References

Abrahams, R. (1998). *Vigilant Citizens: Vigilantism and the State.* Cambridge: Polity.

Byrne, D. N. (2013). 419 digilantes and the frontier of radical justice online. *Radical History Review*, 2013(117), 70 – 82.

Chalfaouat, A. (2020). Empowerment, social distrust or co-production of security: A case study of digital vigilantism in Morocco. In D. Trottier, R. Gabdulhakov, & Q. Huang (eds.), *Introducing Vigilant Audiences* (pp. 161 – 186). Cambridge: Open Book Publishers.

Chang, L. Y., & 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), 1912 – 1932.

Dekker, R., & Meijer, A. (2020). Citizens as aides or adversaries? Police responses to digital vigilantism. In D. Trottier, R. Gabdulhakov, & Q. Huang (eds.), *Introducing Vigilant Audiences* (pp. 281 – 305). Cambridge: Open Book Publishers.

Dennis, K. (2008). Viewpoint: Keeping a close watch – The rise of self-surveillance and the threat of digital exposure. *The Sociological Review*, 56(3), 347 – 357.

Douglas, D. M. (2020). Doxing as audience vigilantism against hate speech. In D. Trottier, R. Gabdulhakov, & Q. Huang (eds.), *Introducing Vigilant Audiences* (pp. 259 – 279). Cambridge: Open Book Publishers.

Favarel-Garrigues, G. (2020). Digital vigilantism and anti-paedophile activism in Russia. Between civic involvement in law enforcement, moral policing and business venture. *Global Crime*, 21(3 – 4), 306 – 326.

Gabdulhakov, R. (2019). In the bullseye of vigilantes: Mediated vulnerabilities of Kyrgyz labour migrants in Russia. *Media and Communication*, 7(2), 230 – 241.

Gabdulhakov, R. (2020). Heroes or hooligans? Media portrayal of StopXam (Stop a Douchebag) vigilantes in Russia. *Laboratorium*, 11(3), 16 – 45.

Gainford, A. L. (2022). *The Challenges of Digilantism for Law Enforcement and Public Order* (Doctoral dissertation, University of Leeds).

Galleguillos, S. (2022). Digilantism, discrimination, and punitive attitudes: A digital vigilantism model. *Crime, Media, Culture*, 18(3), 353 – 374.

Hadjimatheou, K. (2021). Citizen-led digital policing and democratic norms: The case of self-styled paedophile hunters. *Criminology & Criminal Justice*, 21(4), 547 – 565.

Huang, Q. (2023). The discursive construction of populist and misogynist nationalism: Digital vigilantism against unpatriotic intellectual women in China. *Social Media+ Society*, 9(2). https://doi.org/10.1177/20563051231170816

Huey, L., Nhan, J., & Broll, R. (2013). 'Uppity civilians' and 'cyber-vigilantes': The role of the general public in policing cyber-crime. *Criminology & Criminal Justice*, 13(1), 81 – 97.

Jane, E. A. (2016). Online misogyny and feminist digilantism. *Continuum*, 30(3), 284 – 297.

Jane, E. A. (2017). “Dude … stop the spread”: Antagonism, agonism, and #manspreading on social media. *International Journal of Cultural Studies*, 20(5), 459 – 475.

Johnston, L. (1996). What is vigilantism? *The British Journal of Criminology*, 36(2), 220 – 236.

Loveluck, B. (2020). The many shades of digital vigilantism. A typology of online self-justice. *Global Crime*, 21(3 – 4), 213 – 241.

Madsen, M. L. (2004). Living for home: Policing immorality among undocumented migrants in Johannesburg. *African Studies*, 63(2), 173 – 192.

Mano, W. (2010). Between citizen and vigilante journalism: ZimDaily’s Fair Deal Campaign and the Zimbabwe crisis. *Communicare: Journal for Communication Sciences in Southern Africa*, 29(sed-1), 57 – 70.

Milbrandt, T. (2020). ‘Make them famous’: Digital vigilantism and virtuous denunciation after Charlottesville. In D. Trottier, R. Gabdulhakov, & Q. Huang (eds.), *Introducing Vigilant Audiences* (pp. 215 – 258). Cambridge: Open Book Publishers.

Myles, D., Benoit-Barné, C., & Millerand, F. (2020). ‘Not your personal army!’ Investigating the organizing property of retributive vigilantism in a Reddit collective of websleuths. *Information, Communication & Society*, 23(3), 317 – 336.

Ng, E. (2022). *Cancel Culture: A Critical Introduction.* Cham: Palgrave Macmillan.

Plesničar, M. M., & Šarf, P. (2020). ‘This web page should not exist’: A case study of online shaming in Slovenia. In D. Trottier, R. Gabdulhakov, & Q. Huang (eds.), *Introducing Vigilant Audiences* (pp. 187 – 213). Cambridge: Open Book Publishers,

Reichl, F. (2019). From vigilantism to digilantism? In B. Akhgar, P. S. Bayerl, & G. Leventakis (eds.), *Social Media Strategy in Policing: From Cultural Intelligence to Community Policing* (pp. 117 – 138). Cham: Springer International Publishing.

Rosenbaum, H. J., & Sederberg, P. C. (1974). Vigilantism: An analysis of establishment violence. *Comparative Politics*, 6(4), 541 – 570.

Schwarz, K. C., & Richey, L. A. (2019). Humanitarian humor, digilantism, and the dilemmas of representing volunteer tourism on social media. *New Media & Society*, 21(9), 1928 – 1946.

Shearing, C., & Wood, J. (2003). Nodal governance, democracy, and the new ‘denizens’. *Journal of Law and Society*, 30(3), 400 – 419.

Solove, D. J. (2007). *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet.* New Haven: Yale University Press.

Tanner, S., & Campana, A. (2020). “Watchful citizens” and digital vigilantism: A case study of the far right in Quebec. *Global Crime*, 21(3 – 4), 262 – 282.

Trottier, D. (2017). Digital vigilantism as weaponisation of visibility. *Philosophy & Technology*, 30, 55 – 72.

Trottier, D. (2020). Denunciation and doxing: Towards a conceptual model of digital vigilantism. *Global Crime*, 21(3 – 4), 196 – 212.

Vicenová, R. (2020). The role of digital media in the strategies of far-right vigilante groups in Slovakia. *Global Crime*, 21(3 – 4), 242 – 261.

Williams, A. (2020). Black memes matter: #LivingWhileBlack with Becky and Karen. *Social Media + Society*, 6(4). https://doi.org/10.1177/2056305120981047.

Jacqueline D. Wernimont

23 Digital

Abstract: The keyword entry for 'Digital' explores both the etymological and disciplinary contexts for a term that has been central to the formation of a field of Digital Criminology. Considerations include the relationship of 'digital' to 'cyber' and 21st-century critiques of oversimplifications and assumptions about how 'the digital' operates in individual lives and broader cultural communities.

Keywords: digital, cyber-, security, communities of practice

While 21st-century invocations of 'the digital' tend to point toward computation and electronic devices, prior to the middle of the 20th century the term digital was intimately connected to the human hand. Borrowing from the classical Latin *digitalis*, meaning a length equal to a finger's breadth, post-classical Latin and medieval regional languages used 'digital' to refer to a particular finger or to finger-like measurements. The related noun, 'digit,' appeared in classical Latin as *digitus* and was used in medieval Germanic and romance languages to refer to a whole number less than or equal to ten and, anatomically, to individual fingers or toes.

While the meaning of 'digit' has remained relatively stable and in use in mathematical and anatomical lexicons, the meaning of 'digital' shifted dramatically in English during the 20th century. As an adjective, 'digital' is important in signal and information storage and processing, where it refers to data represented by a series of discrete values, commonly 0 and 1. This usage continued with the development of computing and calculating machines that operated on data as discrete numbers and expanded to refer to many kinds of electronic technology that use data stored and processed as discrete numbers.

In 21st-century usage we tend to see 'digital' as an adjective modifying generalized nouns: digital photography, digital technologies, digital society/culture, digital scholarship, or digital crime. In these contexts, digital often connotes not only electronic but also more modern sociotechnical formations that are often linked to the widespread use of computation and internet connectivity (see Computation by Mazzilli Daechsel). The 2021 European Union Serious and Organized Crime Threat Assessment is deeply concerned with digital practices and technologies and opens by asserting "criminals are digital natives … virtually all criminal activities now feature some online component" (Europol, 2021).

Criminologists are particularly well-positioned to be able to define cyber and digital crime and crime control in terms of harm, ethics, historical socio-technical contexts (see Cybercrime by Holt and Holt). While most international academic institutions and research have addressed digital crime in terms of cybersecurity and located those inquiries within science, technology, engineering, and mathematics or in law programs, Brian Payne and Lora Hadzhidimova (2018) observe that criminal justice research and

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-024

interventions provide important tools for understanding how human behavior shapes our approaches to justice, surveillance, and crime in our highly connected digital cultures.

Creepers and continuities

Security and privacy concerns emerged alongside early computing research and the increasing adoption of computing for commerce and, eventually, personal use. In 1972, the U.S. National Bureau of Standards (NBS) established a Computer Security Program. A year before, the first self-replicating computing program, known as Creeper, was developed by Robert Thomas in 1971 and was quickly followed by Ray Tomlinson's Reaper program, which was designed to chase and delete Creeper (Chen and Robert, 2004). Robert Morris Jr.'s infamous 1988 Morris Worm created the first known Denial of Service, or DoS, attack and disrupted roughly 10 % of the internet at the time. Despite both its impact and Morris' eventual prosecution under the 1986 Computer Fraud and Abuse Act, the Morris Worm was designed to measure the internet, not crash it. Like Creeper and Reaper, Morris Worm was not designed with criminal intent. But by the 1990s we saw an explosion of new malicious worms, viruses, and other intrusions that were created to gather personal information, access sensitive information, and for financial gain. The rise of the prefix 'cyber-' essentially follows the historical trajectory of computing-specific security, and cybersecurity curricula have been available for academic and professional education since at least the early-1980s (Parrish et al., 2018).

While harm and safety in cyberspace have been concerns for roughly fifty years, as a term 'digital' is a much older and more capacious than 'cyber-'. I will take a moment here to further unpack the term 'digital' so that we can understand its relationships to criminology before returning to security and surveillance.

A history of 'digital'

The discipline of signal processing distinguishes between digital and analog technology in terms of the duration of a signal; digital technology depends on non-continuous (discrete) signals whereas analog technology depends on continuous or wave signals. These analog signals have an infinite set of possible values, whereas digital signaling depends on a finite set of values. Examples of analog communication include landline/copper wire telephony, magnetic tapes, and classic microphones. In contrast, digital information transmission is discrete and often binary, as in the case of our familiar personal computers, in which data is encoded in 0s and 1s. But early computers like ENIAC and measurement tools like the Geiger counter use decade or base-ten encodings.

While common usage tends to conflate digital information with/in electronic devices, we have had and continue to have digital information outside of electronics. Non-electronic forms of digital information include the abacus (it's also a computer), many

kinds of weaving and needlework patterns, and punched cards like those used for Bouchon and Jacquard mechanical weaving, Carpentier's melotrope music player, Hollerith's tabulating machines used for 19th-century vital statistics, and a variety of accounting machines worldwide. More modern examples of non-electronic digital information include many kinds of student testing technologies where one fills in bubbles on long cards, known as Scantron in the US, and paper voting ballets.

As we have seen, digital communication privileges the discrete and modular whereas analog signals literally make meaning through continuity. As Tara McPherson reminds us, this is a matter of both technical precision and a window into the kinds of cultural logics that digital information shares with other cultural formations (McPherson, 2012). In particular, the technical logics of digital information processing and storage are intertwined with political, social, and knowledge organizations ascendant after World War II. This includes but is not limited to late capitalism, genetics and eugenics, racializing sciences, binary understandings of gender and sexuality, universal knowledge projects, surveillance and policing practices, global migration, and modular labor practices (Chandler and Fuchs, 2019; Chun, 2021; McPherson, 2012; Abbate, 2012; Hicks, 2018; Johnson, 2018; Browne, 2015; Leurs and Smets, 2018; Gajjala, 2019; Woodcock, 2021; Richardson et al., 2019).

Imagining 'digital'

Technical practice and historical usage clearly enable the use of the term 'digital' well outside of either electronics or computing. When Europol describes 21st-century criminals as 'digital natives,' they are leveraging now well-worn rhetoric that imagines generations born into fluency with the "digital language of computers, video games and the Internet" (Prensky cited in boyd, 2014). These imaginaries depend on a slippage between digital information, rendered in discrete signals, and the electronics that have been shaping our lives since the middle of the 20th century.

There are significant socio-technical consequences of these kinds of slippages; as danah boyd so trenchantly observes, applying the term 'digital native' based on birthdate obscures uneven distributions of technological skills, media literacy, and infrastructure access within nations and across the globe (boyd, 2014). It also assumes that people born after a certain date possess uniform cultural and legal literacy when engaging across a wide variety of socio-technical contexts, including cell phone use, activity tracker wearing, biomedical device use, online research, internet banking, and even basic job and school applications. The implications are not just about access, but also about inequities in safety and security when access is possible. Digital harassment, revealing personal information without consent, and stalking are just a few of the many harms that women, people of color, queer people, and people from religious and national minorities experience when trying to engage in activities of daily living that depend on digital technologies.

In addition to problematizing the notion that young people are somehow more fluent and skilled in their engagements with digital information machines, networks, and processes, feminist scholars ask us to attend to the many ways that design choices and social narratives around 'the digital' have rendered it as something non-material. Cloud computing is a particularly salient example; the metaphor suggests an ariel, immaterial, and potentially infinite space where we can store our digital data, rather than a terrestrial network of server stacks residing in electricity and water hungry corporate warehouses. Similarly, both the internet and world wide web may seem abstract to those not familiar with the subterranean, submarine, and satellite infrastructures that connect a series of physical computers to create that web or net. Nicole Starosielski's work on submarine cables is just one example of the kind of recent scholarship that illuminates not only the physical infrastructures essential to digital life, but also their environmental impacts and roles in geopolitical power struggles (Starosielski, 2015).

Indeed, for many our modern computers and computational processes function as 'black boxes,' opaque technical systems where inputs and outputs are visible but the operations between are hidden from user and regulatory view. Unlike transparent 'glass box' operations, black boxes not only hide the technical operations within a system, but they also obscure the material, affective, embodied, labor, and situational contexts that make digital processing and information possible (Wernimont and Losh, 2018). Consequently, most people are insufficiently informed about the possible harms and affordances of digital logics and technologies when it comes time to make behavioral and policy choices.

While common usage might apply 'digital' in an overly broad manner with respect to technical usage, it is also the case that many 21st-century practices and interactions are thoroughly co-constructed alongside our epistemic dependence on discrete information and technology regimes. Consequently, we can and should understand modern 'digital' culture as part of a longer era of discrete, modular, and categorical ways of understanding the world and our societies. In many ways what we now call the digital age exists inside of an era of 'datafication' that extends back to 17th-century vital statistics and manual counting machines like the earliest pedometers and odometers (Wernimont, 2019; see also Dijck, 2014).

Digital justice and crime

Criminology and justice studies seek to understand how socio-technical contexts and human factors shape what a given culture names crime and justice, often expanding well beyond the literal letter of the law or legal precedent. In this context, the adoption of the phrase 'digital criminology' helpfully demarcates different boundaries from cybersecurity and cybercrime, both of which tend to focus on the limited domain of the computer or networks of computers and often emphasize technical processes, defenses, and resolutions (see Parrish et al., 2018; Wirkuttis and Klein, 2017).

Digital criminology is also distinctive in that it draws on longer histories of data-based crime, victimization, and justice. It offers us ways of thinking about guardianship of digital infrastructure, which can extend well beyond the computational, as well as of the many people and flows of products, services, and information in digital cultures (see Infrastructure by Grisot and Parmiggiani). As Smith et al. (2017) observe, digital criminology is "concerned with exploring the critical role performed by digital devices/infrastructures/data as intersectional and transformative mediums in the criminal justice field," as well as reckoning with the perils and possibilities introduced when digital tools are adopted in surveillance, policing, adjudicating, and determining what constitutes justice for those harmed (p. 272).

At the same time, we do well to remember the unequal distributions of power, ability, and harm embedded in modern digital life. Scholars have already enumerated a long list of irregularities and discriminatory outcomes in predictive policing and use of deeply racist or otherwise biased historical data sets (see Prediction by Gundhus, Galis, and Ķīlis; and Policing by Wilson); ethical and labor concerns in the use of synthetic images in time-based media; political and legal consequences of generated image, video, and audio recording; and intellectual property rights for scientific knowledge, artistic creation, and technical innovation (for examples see Browning and Arrigo, 2021; Egbert, 2019; Keyes et al., 2019; Pawelec, 2022; Smits and Borghuis, 2022).

Some of the harms we can point to today build on longstanding concerns about how different actors understand intellectual property, illicit content, and national security (to name just a few areas). But there is also a legacy of colonialism and imbalances in global power that underpins both the way people tend to talk about 'the digital' and people's access to digital resources and opportunities. The digital divide is very real; in 2023, 2.7 billion people (1/3 of the global population) did not have access to the internet and 53% of the world did not have access to high-speed broadband (Signe, 2023). Technological infrastructure of the sort that is now more common than not in wealthy Western nations, is clearly not the only way people can relate to one another or engage in the business of life, nor should it be. Given the psychological, social, political, and environmental costs of those resources, it is not a given that societies who do not use digital infrastructures should want to have it or must be convinced of its value so that they can become the 'next billion users' (Pichai, 2015). Nations and people must also be able to exercise self-determination in relation to socio-technical structures, rather than simply be transformed into new products and users by 21st-century mega-corporations operating as new colonial powers.

At the same time, understanding how harms are perpetuated through digital infrastructure and infostructures is necessary for protecting our most vulnerable people. To offer just a couple of examples: We need better articulations of the harms and possible remedies for women whose pregnancies are revealed without their consent through mundane technologies like loyalty programs or digital temperature monitors. Similarly, though at a different scale, we must articulate how the disruption of digital communication networks impacts survival chances of civilians caught in war. Finally, we are confronted with a double-edged sword in the proliferation of surveillance through

ubiquitous computing. The experience of perpetual risk for some (Black and Brown communities in the United States, for example) has its other in the potential to document war crimes and crimes against humanity for others.

Looking forward

In 1976 the United Nations affirmed that human rights include a right to self-determination and the ability to freely pursue economic, social, and cultural development (International Covenant on Economic, Social and Cultural Rights). The history of international human rights law and of our modern computing cultures and technologies have been deeply intertwined, even if that entanglement has not always been visible. The interdisciplinary field of critical digital justice and criminology can help 21st-century scholars and practitioners maintain the focus on the interplay of sociotechnical systems and situated practices in fostering more just digital societies, more self-determination with respect to digital practices and tools, and our shared understandings of how and why digital crimes occur. These are important interventions in contexts that historically have been dominated by discussions of cybercrime and national-level security concerns. As scholars like danah boyd and others have reminded us, the notion of digital natives is fundamentally flawed even as we find so many of our lives thoroughly mediated by digital technologies. Digital and critical criminologists are essential to our efforts to develop more just futures for not just the well-connected (either in terms of politics or technology) and wealthy but also for people across the world who are seeking more just futures that include reparative modes of dealing with harms perpetuated across the spectrum from interpersonal to international. As Powell et al. (2018) note, this includes essential questions about justice: “What does it mean to be a ’just’ digital citizen? How will digital communications and social networks enable new forms of justice and justice movements?” Digital criminologists need to help us all ask and answer these questions.

Main takeaways

- The history of the term ‘digital’ extends back to antiquity where it referred to an embodied measurement.
- Digital information encoding utilizes discrete values and modular logics and can take both electronic and non-electronic forms.
- Safety and security concerns for digital technologies are as old as the technologies themselves.
- ‘Digital Criminology’ refers to a more capacious field that deals with sociotechnical cultural systems rather than the narrower technical emphasis of cyber-disciplines.

– Care must be taken when using 'digital' because the use of the term has often been a way to obscure material, embodied, and affective considerations for both digital infrastructure and information.

Suggested reading

Brown, M., & Carrabine, E. (2019). The critical foundations of visual criminology: The state, crisis, and the sensory. *Critical Criminology*, 27, 191 – 205.

Brown, S. (2006). The criminology of hybrids: Rethinking crime and law in technosocial networks. *Theoretical Criminology*, 10(2), 223 – 244.

Gordon, F., McGovern, A., Thompson, C., & Wood, M. A. (2022). Beyond cybercrime: New perspectives on crime, harm and digital technologies. *International Journal for Crime, Justice and Social Democracy*, 11(1), i–viii.

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society.* London: Routledge.

Schuilenburg, M., & Peeters, R. (eds.) (2020). *The Algorithmic Society: Technology, Power, and Knowledge.* London: Routledge.

References

Abbate, J. (2012). *Recoding Gender: Women's Changing Participation in Computing.* Cambridge, MA: MIT Press.

boyd, d. (2014). *It's Complicated: The Social Lives of Networked Teens.* New Haven: Yale University Press. https://doi.org/10.12987/9780300166439

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness.* Durham, NC: Duke University Press.

Browning, M., & Arrigo, B. (2021). Stop and risk: Policing, data, and the digital age of discrimination. *American Journal of Criminal Justice*, 46, 298 – 316.

Chandler, D., & Fuchs, C. (2019). *Digital Objects, Digital Subjects: Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data.* London: University of Westminster Press.

Chen, T., & Robert, J.-M. (2004). The evolution of viruses and worms. In W. W. S. Chen, *Statistical Methods in Computer Security* (chapter 16). Boca Raton: CRC Press. 10.1201/9781420030884.ch16.

Chun, W. (2021). *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition,* Cambridge, MA: MIT Press.

Dijck, J. V. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance and Society*, 12, 197 – 208.

Egbert, S. (2019). Predictive policing and the platformization of police work. *Surveillance & Society*, 17(1/2), 83 – 88.

Europol. (2021). *European Union Serious and Organized Crime Threat Assessment.* https://www.europol.europa.eu/cms/sites/default/files/documents/socta2021_1.pdf

Gajjala, R. (2019). *Digital Diasporas: Labor, Affect and Technomediation in South Asia.* Lexington Press, UK.

Hicks, M. (2018). *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing.* Cambridge, MA: MIT Press.

Johnson, J. M. (2018). Markup bodies: Black [life] studies and slavery [death] studies at the digital crossroads. *Social Text*, 36, 57 – 79. https://doi.org/10.1215/01642472-7145658

Keyes, O., Stevens, N., & Wernimont, J. (2019). *The Government is Using the Most Vulnerable People to Test Facial Recognition Software.* Slate.com. https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html

Leurs, K., & Smets, K. (2018). Five questions for digital migration studies: Learning from digital connectivity and forced migration in(to) Europe. *Social Media + Society*, 4(1). https://doi.org/10.1177/2056305118764425

McPherson, T. (2012). Why are the digital humanities so white? Or thinking the histories of race and computation. In M. K Gold (ed.), *Debates in the Digital Humanities* (pp. 139 – 160). Minneapolis: University of Minnesota Press.

Parrish, A., Impagliazzo, J., Raj, R. K., Santos, H., Asghar, M. R., Jøsang, A., Pereira, T., & Stavrou, E. (2018). Global perspectives on cybersecurity education for 2030: A case for a meta-discipline. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (ITiCSE 2018 Companion) (pp. 36 – 54). Association for Computing Machinery, New York, NY, USA. https://doi-org.dartmouth.idm.oclc.org/10.1145/3293881.3295778

Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *DISO*, 1, 19. https://doi.org/10.1007/s44206-022-00010 – 6

Payne, B. K., & Hadzhidimova, L. (2018). Cyber security and criminal justice programs in the United States: Exploring the intersections. *International Journal of Criminal Justice Sciences*, 13(2), 385 – 404. doi:10.5281/zenodo.2657646

Pichai, S. (2015). Google I/O 2015 Keynote. Google also rolled out a section of their blog entitled "Next Billion Users". https://blog.google/technology/next-billion-users/

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society.* London: Routledge.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online*, 94, 15 – 57.

Signe, L. (2023). *Fixing the global digital divide and digital access gap.* Brookings Online. https://www.brookings.edu/articles/fixing-the-global-digital-divide-and-digital-access-gap/

Smith, G., Bennett Moses, L., & Chan, J. (2017). The challenges of doing criminology in the big data era: Towards a digital and data-driven approach. *The British Journal of Criminology*, 57(2), 259 – 274,

Smits, J., & Borghuis, T. (2022). Generative AI and intellectual property rights. In B. Custers & E. Fosch-Villaronga (eds.), *Law and Artificial Intelligence.* Information Technology and Law Series, 35 (pp. 323 – 344). The Hague: T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-523-2_17

Starosielski, N. (2015). *The Undersea Network (Sign, Storage, Transmission).* Durham, NC: Duke University Press.

Wernimont, J. (2019). *Numbered Lives: Life and Death in Quantum Media.* Cambridge, MA: The MIT Press.

Wernimont, J., & Losh, L. (2018). *Bodies of Information: Intersectional Feminism and Digital Humanities.* Minneapolis, MN: University of Minnesota Press.

Woodcock, J. (2021). *The Fight Against Platform Capitalism: An Inquiry into the Global Struggles of the Gig Economy.* London: University of Westminster Press. https://doi.org/10.16997/book51

Wirkuttis, N., & Klein, H. (2017). Artificial intelligence in cybersecurity. *Cyber, Intelligence, and Security*, 1(1), 103 – 119.

Mareile Kaufmann

24 DNA / big genome data

Abstract: DNA has gone big. Not only is our genome increasingly subject to surveillance, but DNA data are also analyzed with ever-changing digital instruments. Genomic analyses and predictions have become inseparable from sophisticated technologies that produce knowledge in criminologically relevant arenas including anything from prevention to forensic lab work and justice. Taking its vantage point in early criminological practices of measuring and counting, this chapter traces how DNA has risen to become a central type of biodata in today's surveillance societies.

Keywords: surveillance, police, forensics, predictions, biometrics

Our genomic material is increasingly registered and digitized. As Erin Murphy puts it: "Big genome data has arrived" (2018: e8). DNA is sequenced with ever-new techniques, stored in numerous databases, and processed with algorithms (see Algorithm by Leese). DNA data and the results of their analyses circulate in public and private domains, from medicine to ancestry research, but not least in the field of law enforcement. Studying the integration of digital technologies with genomics is relevant to criminologists because it influences forensic, police, and courtroom practices as well as raising issues of surveilling bodies and bio-data. What is more, the criminological engagement with these themes can be considered essential, because measuring bodies, collecting biometric data, and processing them for investigative purposes is intricately linked to criminology as a discipline.

Indeed, studying the use of DNA technologies in law enforcement is nothing peculiar in light of the early criminologists' project to scientize the discipline by measuring bodies and translating them into statistical systems. Throughout the 19th century a plethora of practices emerged that sought to metricize human bodies, many of them in the field of criminalistics. It was the rise of photography, but also instruments such as thermometers, dynanometers for force and weight, and stereographs for profiling skulls that sparked the development of metric systems, that is diverse measuring practices, for the analysis and categorization of humans. Cesare Lombroso's (1835–1909) infamous project of classifying criminal bodies by type comes to mind. However, contem-

Disclaimers and acknowledgments: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 947681).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-025

poraries such as Paul Broca (1824–1880), Francis Galton (1822–1911), Alphonse Bertillon (1853–1914) and Edmond Locard (1877–1966) were contributors to the anthropometric and taxonomic project at the time. Controversies over the 'correct' metric system, method, and classification kept emerging (see Gibson and Rafter, 2006). These controversies illustrate a key point about the use of biometrics in law enforcement at large: technology is not neutral—it is always tied to the design, underlying assumptions, and use by humans.

Today, the use of biometry for forensics, such as DNA profiling, follows much stricter protocol and forensic scientists are aware of the probabilistic nature of the knowledge they produce. And still, the ways in which genomic data travel across databases and are analyzed by software is highly dependent on the humans interacting with such technologies (see Databases by Bellanova).

The rise of DNA

In order to understand the ways in which DNA is relevant to law enforcement, it is necessary to understand some basics. The roots of DNA analysis lie in bio-chemistry and blood typing, but also in hereditary studies that emerged in the mid-19th century. The 'discovery' of DNA and its integration into scientific research was not the effort of a single scientist. Johann Friedrich Miescher discovered nucleic acids in 1869, which would later be termed DNA; in the mid 1900s Phoebus Levene identified several components that led to an understanding of DNA's functioning and Erwin Chargaff uncovered details of the structure of DNA (Pray, 2008). Building on the work of many others, James Watson and Francis Crick finally identified the double-stranded helix of DNA (1953). Another momentous finding was that repeating sequences on the genome could be 'informative markers' and unique to each individual (Jeffreys et al., 1985). Some of these markers would be placed in the non-coding regions of the genome. While the coding regions of the genome encode proteins for different bodily functions and traits, non-coding regions vary much more between individuals which is why they are suited for identification. Such variations show in short sequences, so-called Short Tandem Repeats (STRs). While early methods revealed four STRs, later sequencing procedures revealed 18, then 24/25 STRs. Early sequencing techniques visualized STRs as barcodes on membranes, which then had to be compared in an analog fashion. With the digitization of such data, DNA profiles could be more efficiently stored and exchanged across laboratories, police databases, and national borders. This means that DNA profiles could be more easily compared to identify matches between a DNA profile in a database and that of an unknown owner.

The use of DNA in law enforcement

DNA matching was long contested as a reliable source of identification. Comparisons between DNA profiles, for example, never result in a 1:1 match. Instead, they work with approximations and DNA matches are expressed as probabilities. Michael Lynch et al. (2008) have traced the 'DNA wars' of the 1980s and 1990s, which revolved around controversies on sample collection and sampling methods, data availability and probabilistic modelling. It was thus necessary to develop protocol. The US National Research Council, for example, notes that such differences in technique can be documented in a laboratory protocol that "should not only specify steps and reagents, but also provide precise instructions for interpreting results, which is crucial for evaluating the reliability of a method" (NRC, 1992: 53).

Protocols or standard operating procedures, however, are enacted differently in practice (Lynch et al., 2008). They are appended, interpreted, and refined and many of such instances remain tacit—a known phenomenon amongst forensic staff (Lynch et al., 2008: 104). That is to say, even at the level of DNA matching, uncertainties exist, which is also expressed by the probabilistic communication of DNA 'matches.' That is, capturing and communicating these uncertainties is also a socio-technological practice.

By today, Next Generation Sequencing (NGS) established itself as an alternative to classic STR sequencing. NGS is a "high throughput method" utilizing "DNA sequencing technologies that are capable of processing multiple DNA sequences in parallel" (National Cancer Institute, n.d.). The result is a more comprehensive sequence that also allows for the discovery of so-called Single Nucleotide Polymorphisms (SNPs), which are genetic markers for specific traits. This can be useful, for example, in medicine for identifying a genetic disposition for a specific disease. Since the 2010s the method also gained traction in the forensic sciences and sparks growing interest. For example, NGS gives better results for degraded DNA samples (Lackey, 2017) or makes it easier to "distinguish the different DNA profiles in a complex mixture" (Bruijns et al., 2015: 2647). The latter is important, because most traces contain DNA from several owners and such mixtures are still a major challenge for forensic science. The more sequenced material available, the higher the likelihood of being able to distinguish DNA owners from each other. What is more, SNPs can be used as predictors for analyzing observable traits of an individual when a DNA match cannot be established. This process is called phenotyping. DNA profiles based on NGS can also be used for familial searches, which is to look for genetic relatedness to identify relatives of a potential suspect. Both will be discussed below.

While NGS is already an accepted sequencing technique in the global north, opinions about using DNA profiles for more than matching is divided among the police and forensic communities (Samuel and Prainsack, 2018). While some practitioners and product developers consider the predictive powers of such techniques a major progress in the field of forensics, others question the usability of their results and point to eth-

ical issues. Some countries, like the US, France, the UK, and the Netherlands have already integrated phenotyping and familial searches into forensic practice at the time of writing. Differences exist in terms of what type of phenotypic information can be used; especially phenotyping that can reveal medical conditions is more strictly regulated. Other countries have not yet adopted any of these techniques which could be for fiscal, practical, legal, or ethical reasons. Judging by the rising number of publications, products, and conference tracks about phenotyping and familial searches, however, an increase in their deployment is expectable.

The relationship between DNA and digital technologies: key phenomena

This entry fronts three phenomena as relevant research areas for criminologists. With the increasing integration of DNA and digital technologies these areas will keep on gaining traction. The most remarkable shift is the move from DNA as a tool for identification to a tool for prediction and investigation.

Databases

DNA profiles need storage in order to function as a source of identification. DNA data can now be stored in electronic databases, they can be analyzed with digital tools and exchanged between laboratories or police institutions. This leads to a rapid growth of DNA databases with many of them being, however, decentral. In the US, for example, such databases are run by different states, different public institutions, or commercial providers, because they grew out of different local forensic traditions. Here, we find many variations in information practices that influence the quality of the database: "It is flawed human beings that collect their information, write their operating codes, input their entries, maintain their systems, and search and retrieve their data. Databases may represent turbo-charged knowledge—but it is still human knowledge, just more powerful" (Murphy, 2010b: 825).

In Europe, the PRÜM treaty was created in 2005 to establish a more streamlined system for DNA databases across Europe, even requiring member states to set up DNA databases for information exchange (European Data Protection Supervisor, 2007, no page number). However, while centralized working groups and competence teams for support were established by the EU, the different stages of database development, problems with the availability and exchange of information, but also high costs accompany the implementation of the system (Jones, 2012).

Since DNA databases also form the fundament for algorithmic analysis, the composition of stored data moves to the fore. Here, scholars have voiced concern, for example, over the categorization and overrepresentation of minorities in such databases

(Murphy and Tong, 2020; Skinner 2020). The ways in which DNA databanks are organized, how and whether they speak to each other, need continued scholarly engagement in order to understand the role of digital DNA systems in law enforcement.

Phenotyping

A phenotype relates to an individual's observable features. In genetic phenotyping, the abovementioned SNPs work as predictors for analyzing traits of an individual. This procedure can amongst other things be used for medical, but also for forensic purposes. For the latter traits such as hair-, eye- and skin-color are of interest. Phenotyping efforts can also be directed at predicting age or geographical heritage by continental or sub-regional type (Kayser et al., 2023). When used for law enforcement, phenotyping is supposed to assist in investigations, especially when a DNA match cannot be established.

For phenotyping one needs large data pools that allow for a comparison of genes known to impact, for example, pigmentation with data of individuals whose geno- and phenotype is already known. An algorithm infers from these datasets which genetic marker is likely to influence what type of pigmentation. In order to obtain a graded (i. e., not continuous) visualization within the software, pigmentation is categorized. This procedure, as Roos Hopman criticizes, reflects everyday racial categories, such as 'white,' 'intermediate,' and 'black' and inscribes them into software (2023). Some phenotyping software solutions, such as Parabon Snapshot®, also seek to predict the shape of the face. For this, facial morphologies are captured and standardized, cleaned of irregularities, symmetrized, and rendered comparable (Hopman, 2023). Similar to the process above, an algorithm infers which genomic markers likely express typical morphological features. All these physiognomic results are of predictive nature: they do not describe an actual person, but predict a face (Wienroth, 2018). While practitioners are mainly concerned about the practicability of such predictions (a person can dye their hair, wear contact lenses or makeup), more fundamental critique is directed at the making and the effects of 'typing.' Indeed, not an individual's face, but a pheno*type* is produced: "The focus shifts from the genetically unique to the genetically common" (Hopman, 2020). The ambition of phenotyping may be accuracy, but such predictions never articulate accuracy, only an appearance of truth that results in categorical predictions (Hopman, 2020). Though the logic of the category was already operative in earlier biometric practices (Cole, 2001), acts of categorizing bodies based on genetic make-up express new generalizations that are now part of identification procedures. Or, as Hopman and M'charek write: "the promise of individuality, namely the face of the suspect, is quintessentially about producing the collective" (2020: 460). Not only is the net for investigation broadened, drawing entire collectives into the realm of suspicion, but the concept of the collective opens up for categorizations based on race (Hopman and M'charek, 2020). The future of phenotyping will see efforts to type

more and different characteristics, the implications of which are important to understand and discuss in detail.

Familial searches

First implemented in the UK in 2002, more and more countries are adopting familial searches, that is biological relatedness of DNA samples. In most countries, however, its use is restricted to serious cases (Granja, 2021). The technique gained traction with the advent of NGS, because the higher number of identified STRs and SNPs enables a search for genetic relatedness. Originally, a matching exercise of two DNA samples would aim to identify an as exact as possible match of 26 alleles. The more alleles match, the higher is the likelihood that two DNA samples stem from the same person. With the rise of familial searches, however, a conscious decision is taken to compare more comprehensive genomic material to focus specifically on genetic mutations or part-matches that can indicate biological relatedness. The technique, then, creates a larger pool of people from which new leads for investigation can be inferred. However, this also means that more individuals are drawn into such investigative processes. As Erin Murphy summarizes it: familial searches do not provide matches but only 'gestures'; they "point only to a list of possible suspects, all but one of whom definitely did not leave the evidence" (2010a: 313). Consequently, familial searches "affect more people, because they cast the net indiscriminately and widely" (Murphy, 2010a).

The increase in publicly owned databases, but also recreational databases, such as *FamilyTreeDNA*, *23andMe*, and *Ancestry* that offer the analysis of an individual's genealogy as a service, add a new dimension to familial searches. In the latter, individuals chose to let their DNA be analyzed to learn more about their genetic 'roots' based on a wider net for genetic overlaps. These are long-range familial searches that are also called 'forensic genealogy' when adopted by law enforcement. The use of recreational databases for forensic purposes enlarges datasets and data subjects that can be used for investigation. At the same time, this means that the "involvement with the criminal justice system is no longer a pre-requisite to participate in law enforcement searches" (Granja, 2021: 339). This and the probabilistic nature of the technique are ethical concerns and produce a clear shift in the use of DNA in law enforcement. A different dimension of this trend is that genetic knowledge is increasingly commodified, creating issues of transparency, ownership, and protection of genomic data.

Conclusion

The integration of genetics with digital technologies produces new phenomena, practices, and politics in the field of law enforcement. While the project of metrifying and typing human bodies is as old as the discipline of criminology, it takes new shape with DNA data 'going big.' For a long time, the main use of DNA was identification,

whereby the likely match between several samples is assessed. Today, the practice of DNA matching is joined by several big data practices (see Big Data by Završnik). Not only do databases grow in size and number, but today's sequencing methods can reveal ever-more comprehensive information as well as genetic markers for specific traits. While such advances generate new avenues for investigation and identification, for example from sample mixtures, they also introduce a double shift: from ascertaining identity towards prediction, and from a focus on individuals to a focus on collectives. New investigative methods such as familial searches or the prediction of body 'types' are symptomatic of this shift. These developments are accompanied by a range of issues relevant to criminologists:

- How are databases created and maintained? Can and should databases communicate with each other?
- How do new software models generate phenotypes and what exactly are they 'typing'? What assumptions and decisions are folded into these models? What are the promises and pitfalls of predicting types altogether?
- Which relations are traced in familial searches and forensic genealogy? What does it mean that genetic relation, history, and geography are now a part of police investigation?
- How are digital tools for DNA analysis used in practice? How do genomic prediction and uncertainty influence a context as sensitive as law enforcement? What are the effects of the drift from the individual to the collective?
- How are forensic technologies spreading? There is literature on the 'professional gaze' in forensic genetics (Granja and Machado, 2019; Samuel and Prainsack, 2018). Yet, how do attitudes change over time?
- Much has been written about the ethical responsibilities and controversies in forensic genetics (e.g., Wienroth et al., 2021, Roux et al., 2022), but what *ethos* do genomic technologies perform?
- What is the role of commercial providers in genetic knowledge-making?
- What are legal and regulatory challenges (cf. Scudder et al., 2018)?

Suggested reading

Granja, R. (2021). Long-range familial searches in recreational DNA databases: Expansion of affected populations, the participatory turn, and the co-production of biovalue. *New Genetics and Society*, 40(3), 331 – 352.

Hopman, R. (2023). The face as folded object: Race and the problems with 'progress' in forensic DNA phenotyping. *Social Studies of Science*, 53(6), 869 – 890. https://doi.org/10.1177/03063127211035562

Murphy, E. (2010). Relative doubt: Familial searches of DNA databases. *Michigan Law Review*, 109(3), 291 – 348.

References

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*, 39, 2642 – 2654.

Cole, S. (2001). *Suspect Identities. A History of Fingerprinting and Criminal Identification.* Cambridge, MA: Harvard University Press.

European Data Protection Supervisor. (2007). *Cross-border cooperation (Prüm Treaty).* https://edps.europa.eu/data-protection/our-work/publications/opinions/cross-border-cooperation-pr%C3%BCm-treaty_en (Accessed 5 September 2022)

Gibson, M., & Rafter, N. (2006). Editors' introduction. In D. Lombroso (ed.), *Criminal Man* (pp. 1 – 35). Durham, NC: Duke University Press.

Granja, R. (2021). Long-range familial searches in recreational DNA databases: Expansion of affected populations, the participatory turn, and the co-production of biovalue. *New Genetics and Society*, 40(3), 331 – 352.

Granja, R., & Machado, H. (2019). Ethical controversies of familial searching: The views of stakeholders in the United Kingdom and in Poland. *Science, Technology, & Human Values*, 44(6), 1068 – 1092.

Hopman, R. (2020). Opening up forensic DNA phenotyping: The logics of accuracy, commonality and valuing. *New Genetics and Society*, 39(4), 424 – 440.

Hopman, R. (2023). The face as folded object: Race and the problems with 'progress' in forensic DNA phenotyping. *Social Studies of Science*, 53(6), 869 – 890. https://doi.org/10.1177/03063127211035562

Hopman, R., & M'charek, A. (2020). Facing the unknown suspect: Forensic DNA phenotyping and the oscillation between the individual and the collective. *BioSocieties*, 15(3), 438 – 462. https://doi.org/10.1057/s41292-020-00190-9.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314, 67 – 73.

Jones, C. (2012). Complex, technologically fraught and expensive' – The problematic implementation of the Prüm Decision. *Statewatch Journal*, 22(1), 1 – 7 https://www.statewatch.org/media/documents/analyses/no-197-prum-implementation.pdf (Accessed 10 November 2023)

Kayser, M., Branicki, W., Parson, W., & Phillips, C. (2023). Recent advances in forensic DNA phenotyping of appearance, ancestry and age. *Forensic Science International: Genetics*, 65: 102870.

Lackey, A. (2017). *Capillary Electrophoresis (CE) or Next-Generation Sequencing (NGS) for STR Analysis?* ThermoFisher Scientific. https://www.thermofisher.com/blog/behindthebench/capillary-electrophoresis-or-next-generation-sequencing-for-str-analysis/ (accessed 9 November 2023)

Lynch, M., Cole, S. A., McNally, R., & Jordan, K. (2008). *Truth Machine. The Contentious History of DNA Fingerprinting.* Chicago: The University of Chicago Press.

Murphy, E. (2010a). Relative doubt: Familial searches of DNA databases. *Michigan Law Review*, 109(3), 291 – 348.

Murphy, E. (2010b). Databases, doctrine, and constitutional criminal procedure. *Fordham Urban Law Journal*, 37(3), 803 – 836.

Murphy, E. (2018). Law and policy oversight of familial searches in recreational genealogy databases. *Forensic Science International*, 292, e5–e9.

Murphy, E., & Tong, J. H. (2020). The racial composition of forensic DNA databases. *California Law Review*, 108, 1847 – 1911.

National Cancer Institute. (n.d.). *Massively Parallel Sequencing.* Genetics Dictionary. Available at: https://www.cancer.gov/publications/dictionaries/genetics-dictionary/expand/M (3 September 2022)

National Research Council. (1992). *DNA Technology in Forensic Science.* Washington, DC: National Academy Press.

Pray, L. (2008). Discovery of DNA structure and function: Watson and Crick. *Nature Education*, 1(1): 100.

Roux, C., Bucht, R., Crispino, F., De Forest, P., Lennard, C., Margot, P., Miranda, M. D., et al. (2022). The Sydney Declaration – Revisiting the essence of forensic science through its fundamental principles. *Forensic Science International*, 332, 111182.

Samuel, G., & Prainsack, B. (2018). Forensic DNA phenotyping in Europe: Views "on the ground" from those who have a professional stake in the technology. *New Genetics and Society*, 38(2), 119 – 141.

Scudder, N., McNevin, D., Kelty, S. F., Walsh, S. J., & Robertson, J. (2018). Massively parallel sequencing and the emergence of forensic genomics: Defining the policy and legal issues for law enforcement. *Science & Justice: Journal of the Forensic Science Society*, 58(2), 153 – 158.

Skinner, D. (2020). Race, racism and identification in the era of technosecurity. *Science as Culture*, 29(1), 77 – 99.

Watson, J. D., & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature*, 171, 737 – 738.

Wienroth, M. (2018). Governing anticipatory technology practices. Forensic DNA phenotyping and the forensic genetics community in Europe. *New Genetics and Society*, 37(2), 137 – 152.

Wienroth, M., Granja, R., Lipphardt, V., Amoako, E. N., & McCartney, C. (2021). Ethics as lived practice. Anticipatory capacity and ethical decision-making in forensic genetics. *Genes*, 12(12), 1868.

Claudia Aradau

25 Error

Abstract: This entry traces some of the transformations of error in the management of suspect populations, the policing of crime, and the taming of the uncertainties and complexities of governing populations digitally. It shows that what counts as error has changed from statistics to machine learning. Moreover, error is enacted differently depending on cultural, social, and political articulations of normality and abnormality.

Keywords: error, governmentality, statistics, machine learning, suspect populations

What is error?

Error has an ambiguous meaning. On the one hand, to err means "to ramble, roam, stray, wander" (Oxford English Dictionary, 2023). On the other, to err is to "go astray" (Oxford English Dictionary, 2023) from the correct path, to deviate. While the first meaning is now obsolete, this ambiguity of error as wandering or deviation has preoccupied philosophers and historians since the entry of error in the 18th-century *Encyclopaedia of Diderot and d'Alembert* (Bringman, 2013 [1755]). Given the multitude of forms errors could take, the *Encyclopaedia* entry proposed to find the cause and origin of error. The search for the cause of errors has continued to trouble philosophers and scientists alike.

For the historian of science Lorraine Daston (2005: 5), "epistemology since the seventeenth century has consisted largely in an elaborate nosology of errors: what their species and varieties are, and how they may best be avoided or cured." The classification, evaluation, and avoidance of errors do not only characterize scientific knowledge. The knowledge required for the government of individuals and populations has been shaped by vigilance against error and a corrective ethos. Drawing lines between the normal and the abnormal, the criminal and the law-abiding, the citizen and the non-citizen is fraught with possibilities of and anxieties about error.

Philosopher of science Giora Hon points to the heterogeneity of error as "a multifarious epistemological phenomenon of great breadth and depth" (Hon, 2008). Given this breadth and depth, error is impossible to define, contain, and eliminate. Therefore,

Funding: This work was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (SECURITY FLOWS, Grant Agreement No. 819213).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-026

as governing actors have attempted to tame error, we need to ask: 'how is error?' rather than 'what is error?'. 'How is error?' is an inquiry into what kinds of errors surface more frequently than others and how they can be ranked so that some require more intense attention and vigilance than others. 'How is error?' allows us to map what counts as error in the government of populations, how errors are contested, and how they are transformed through the use of various technologies, including digital technologies.

This entry traces some of the transformations of error in the management of suspect populations, the policing of crime, and the taming of the uncertainties and complexities of governing populations digitally. It shows how errors have always been produced as technical and political.

Governing suspect populations: from statistics to machine learning

From medical errors to errors of justice and from grammatical errors to biometric errors (see Biometric Failure by Magnet and Din), errors are ubiquitous in social and political life. However, some errors carry more severe consequences than others. If errors can never be entirely eliminated, then they need to be hierarchized, categorized, and minimized so that they don't have too detrimental or destructive effects on knowledge production and social life. The categorization and minimization of error played out differently in the historical entwinement of statistics and the government of populations: from the 'error curve' of measurement to error as a failure of scientific methods (Roberts, 2011: 65).

As statistics was deployed to manage populations and draw distinctions between risky and non-risky individuals (Foucault, 2007), error was the rationale that enabled distinctions between normality and abnormality. The awareness that measurement could always contain errors had led astronomers to discover that errors followed an 'error curve.' In the 19th century, the Belgian mathematician Adolphe Quetelet developed his social physics based on the astronomers' theory of error. The 'average man' as the mean of the normal curve distribution was the equivalent of the mean of measurement error. Subsequently, the mean was assumed to be a correct measurement and, therefore, truthful knowledge. The truth of the average had political effects for the untruth of deviation. As historian of statistics Theodore Porter (1985: 65) has aptly put it: "Quetelet's idealization of the mean as the standard of beauty and goodness implied that all variation—everything exceptional—was to be regarded as flawed, the product of error."

Statistics afforded criminology "its intellectual orientation and recognition by the scientific community" (Beirne 1987: 1145). Statistical analysis of deviation from the 'average' enabled distinctions between "the recalcitrant minority with incorrigible criminal tendencies" (Beirne 1987: 1161) and those whose minimal deviances could

be governed through the "inevitable progress of civilization" (Beirne 1987: 1161). The focus on the 'dangerous classes' in the social physics of deviance was entwined with anxieties about mobility or what historian Simon Cole (2002: 9) has described as the "increasing formal criminalization of mobility." For example, fraud and swindling were seen to be "crimes of mobility." Given 19th-century anxieties about mobility and the "dangerous classes," measuring the body and related statistical calculations promised to identify individuals who otherwise might have escaped the gaze of policing. Here, the measurement promised to identify a person with minimal error. As Cole has explained, "the identification techniques of the time focused on rendering identities stable, on the processes of capturing and fixing identity, for which the freeze-frame served as such an appropriate metaphor" (Cole, 2002: 22). Alongside photographs, French policeman Alphonse Bertillon introduced statistics in the measurement of criminal individuals for the purpose of identification. Limiting deviation in measurement was supposed to yield a true measurement, which could then be compared with other individuals' measurements (Bertillon, 1885).

However, statistical measurement, as undertaken by Bertillon, seemed to have offered only limited accuracy and certainty. Eugenicist and initiator of the statistics of variation Francis Galton castigated Bertillon's method of bodily measurements for its errors (Galton, 1892). Error was here a matter of scientific methods rather than measurement. Statistical variables had to be independent, Galton argued, but the variables Bertillon used were not. The measurements of the human body were co-dependent and could only lead to erroneous conclusions. Bertillon's measurements were plagued both by "theoretical error," namely treating measurements that were "undoubtedly correlated" as independent, and practical error of measurement through "the absence of a sufficiently detailed account of the practical difficulties experienced in its employment" (Galton 1892: 158). Taking measurements of the human body required craft, instruments, and experience. Instead of Bertillon's anthropometry, Galton proposed fingerprinting as an almost error-free method. Fingerprints, he contended, "may therefore be treated without the fear of any sensible error" (Galton, 1892: 167).

This controversy was not decisive in how biometrics and anthropometry were subsequently used for the purposes of identification and policing. Another historian of statistics has argued that the popularity of fingerprints was not due so much to infallibility but "to the striking visual appearance of fingerprints in the courts," a few "dramatically successful cases," and the inattention to non-unique or changeable fingerprints (Stigler, 1995: 860). Nevertheless, the effects fingerprints had in courts depended on the imagination of accuracy and certainty that they appeared to offer to questions of identification.

Since Galton's time, biometrics have expanded to identify criminals and their victims, migrants and citizens, and even to detect spies. In the late 20th century, the polygraph or the use of DNA in courts became such public controversies about errors (on the exceptional status of DNA evidence in courts, see Lynch, 2013). In October 1983, the *New Scientist* reported on the introduction of the polygraph at the Government Communications Headquarters (GCHQ), the UK's signals intelligence agency, whose exis-

tence had only been recently recognized by the government. According to the union at GCHQ, "polygraph screening tests wrongly clear one in four of guilty subjects and wrongly accuse 50 per cent of the totally innocent subjects" (New Scientist, 1983). The introduction of the polygraph led to a decade of labour strikes, also in the wake of the prohibition of unions at GCHQ announced by the Thatcher government in 1984 (Aldrich, 2010, chapter 21).

These controversies over the errors of statistics and biometric measurements draw attention to error not just as a scientific and technical concept but also as a social and cultural one. More recently, the literature on biometrics has situated errors in a social and cultural context. Surveillance studies scholar Shoshana Magnet (2011) has argued that the understanding of biometric error needs to be expanded beyond what the biometric industry renders as technical varieties of error: "failure to enroll" and "failure to capture" rates to "false accept" and "false reject" rates. These are what she has called "technological crashes" (Magnet 2011: 13). However, and more significantly, "biometrics fail to work on particular communities in ways connected to race, class, gender, sexuality, and disability" (Magnet 2011: 7). Magnet has mapped the wide range of biometric errors in processes of identification and verification, "from high rates of unbiometrifiability to the misapplication of statistical technique in the codification of bodily identities and the misunderstanding of cultural trends, even with respect to hairstyle and clothing" (Magnet, 2011: 153).

Errors are inescapable given the assumptions of a 'normal body' inscribed in biometric technologies. What counts as 'normal' is a techno-scientific average and a sociocultural normal imbued with assumptions about abnormality. In the case of India's gigantic Aadhaar programme, which assigns a unique number to residents based on biometric data, the "unbiometrifiability of bodies" has been part of "heated debates about the feasibility of aadhaar" (Jacobsen and Rao, 2018: 29). For example, those whose fingerprints are uncapturable or damaged become 'errors.' The production of biometric errors in the Aadhaar system has far-reaching consequences, as anthropologists Elida K. Jacobsen and Ursula Rao have noted: "The discrimination of those not given an aadhaar number is amplified in a context of interoperability, whereby denial at one access point can lead to chains of exclusion" (Jacobsen and Rao, 2018: 34).

More recently, these concerns about error have been magnified through the use of machine learning algorithms to process biometric data (see Algorithm by Leese). While scholars and activists have drawn attention to errors in training data and bias towards gendered and racialized categories of population, the roll-out of machine learning-powered biometric technologies such as facial recognition has continued apace (Aradau and Blanke, 2021). The National Institute of Standards and Technology (NIST) in the USA explains that "Face recognition accuracy has improved markedly due to the development of new recognition algorithms and approaches" (NIST, 2023). However, error is far from eliminated, as NIST acknowledges the persistence of "increased error rates in applications where photography of faces is difficult or when stringent thresholds must be applied to recognition outcomes to reduce false positives" (NIST, 2023).

With machine learning, error entails a process of optimization or "adjusting machine learning models to achieve the best possible performance within a particular use case" (Aradau and Blanke, 2021: 8). The aim is not to eliminate or even necessarily minimize error but to optimize it. The optimization of error is a political decision about what a specific biometric technology does in the world. As digital culture scholar Nanna Bonde Thylstrup (2021: 197) has formulated the problem of big data errors, error is "a fundamental, and fundamentally political, part of knowledge production." She has drawn attention to computer science vocabularies that integrate error into a continuum of terms such as glitch, anomaly, fault, flaw, etc. (Thylstrup, 2021: 193).

These multiplicities, optimizations, and ambiguities of error mean that problematizing error cannot be done in the name of scientific accuracy, precision, validity, or truth. As the history of biometric technologies from the age of statistics to that of machine learning shows, biometric technologies cannot be perfect, both because they are probabilistic and because they are social, cultural, and geopolitical. Critical engagements with error can open new spaces of contestation and problematization of what counts as suspect populations, risky behaviors, and predictions of criminality (see Prediction by Ķīlis, Gundhus and Galis). To do so, critical engagements with error need to problematize the politics of producing and identifying errors, what counts as acceptable or unacceptable error and how distinctions between error, glitch, bias, or failure are drawn (Aradau and Blanke, 2021; Thylstrup, 2021; Ananny, 2022).

Conclusion: how is error?

Error has shaped not only questions about the accuracy of knowledge production but also about how and with which technologies populations are made governable. Starting from the normal curve, which saw the distribution of measurement or behavior as undesirable if deviant from 'averages,' error was imbricated with social, cultural, and political concerns. As error has been mobilized in scientific and public controversies about when and how biometric technologies are used to govern suspect populations, it is crucial to attend to its forms and transformations. What counts as error has changed from statistics to machine learning. Moreover, what counts as error is enacted differently depending on cultural, social, and political articulations of normality and abnormality. For instance, facial recognition has recently attracted intense attention as its errors are deeply entwined with racialized and gendered classifications of populations (see Facial Recognition by Fussey). Yet, from statistics to machine learning, errors have also changed, as the anxieties of identifying and correcting errors have been supplemented by the ambiguities of optimizing errors.

If errors are inscribed into the digital technologies deployed to govern populations, then social scientists need to address their multiple articulations by asking: 'How is error?' Errors reproduce historical patterns of inequality so that racialized individuals and groups come to be read as 'errors.' Errors enable decisions on the thresholds of probability that make some biometrically recognizable and others not. Errors can in-

fuse scientific and academic controversies over the invention and use of some technologies rather than others. Errors also etch economic rationalities of cost and efficiency into what can be problematized, addressed, or redressed and what will remain unspoken and unaccounted. Mapping the variations and mobilizations of error avoids ceding its political terrain to technical competencies only.

Suggested reading

Ananny, M. (2022). Seeing like an algorithmic error: What are algorithmic mistakes, why do they matter, how might they be public problem? *Yale Journal of Law & Technology*, 24, 342 – 364.

Aradau, C., & Blanke, T. (2021). Algorithmic surveillance and the political life of error. *Journal of the History of Knowledge* 2(1).

Daston, L. (2005). Scientific error and the ethos of belief. *Social Research*, 72(1), 1 – 28.

Thylstrup, N. B. (2021). Error. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.), *Uncertain Archives: Critical Keywords for Big Data* (pp. 191 – 199). Cambridge, MA: MIT Press.

References

Aldrich, R. J. (2010). *GCHQ: The Uncensored Story of Britain's Most Secret Intelligence Agency.* London: Harper Press.

Ananny, M. (2022). Seeing like an algorithmic error: What are algorithmic mistakes, why do they matter, how might they be public problem? *Yale Journal of Law & Technology*, 24, 342 – 364.

Aradau, C., & Blanke, T. (2021). Algorithmic surveillance and the political life of error. *Journal of the History of Knowledge*, 2(1). https://journalhistoryknowledge.org/article/view/11370.

Beirne, P. (1987). Adolphe Quetelet and the origins of positivist criminology. *American Journal of Sociology*, 92(5), 1140 – 1169. https://doi.org/10.1086/228630..

Bertillon, A. (1885). *Identification anthropométrique. Instructions signalétiques.* Ministère de l'Intérieur, Paris: Melun.

Bringman, G. (2013 [1755]). Error. *The Encyclopedia of Diderot & d'Alembert Collaborative Translation Project.* Michigan Publishing. Accessed 3 April. http://hdl.handle.net/2027/spo.did2222.0002.683.

Cole, S. A. (2002). *Suspect Identities. A History of Fingerprinting and Criminal Identification.* Cambridge, MA: Harvard University Press.

Daston, L. (2005). Scientific error and the ethos of belief. *Social Research*, 72(1), 1 – 28.

Foucault, M. (2007). *Security, Territory, Population. Lectures at the Collège de France, 1977 – 78.* Basingstoke: Palgrave.

Galton, F. (1892). *Finger Prints.* London: Macmillan and Co.

Hon, G. (2008). Error: The long neglect, the one-sided view, and a typology. In G. Hon, J. Schickore, & F. Steinle (eds.), *Going Amiss in Experimental Research* (pp. 11 – 26). Dordrecht: Springer.

Jacobsen, E. K., & Rao, U. (2018). The truth of the error. In M. Maguire, U. Rao, & N. Zurawski (eds.), *Bodies as Evidence. Security, Knowledge, and Power* (pp. 24 – 42). Durham, NC: Duke University Press.

Lynch, M. (2013). Science, truth, and forensic cultures: The exceptional legal status of DNA evidence. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(1), 60 – 70. https://doi.org/https://doi.org/10.1016/j.shpsc.2012.09.008.

Magnet, S. A. (2011). *When Biometrics Fail: Gender, Race, and the Technology of Identity.* Durham, NC: Duke University Press.

New Scientist. (1983). Why spies like lie detectors. 20 October 1983, 161.

NIST. (2023). *Face Analysis Technology Evaluation (FATE) Quality.* Accessed 29 September. https://pages.nist.gov/frvt/html/frvt_quality.html.

Oxford English Dictionary. (2023). err, v. Oxford: Oxford University Press.

Porter, T. M. (1985). The mathematics of society: Variation and error in Quetelet's Statistics. *The British Journal for the History of Science*, 18(1), 51–69.

Roberts, J. (2011). *The Necessity of Errors.* London: Verso.

Stigler, S. M. (1995). Galton and identification by fingerprints. *Genetics*, 140(3), 857–860.

Thylstrup, N. B. (2021). Error. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.), *Uncertain Archives: Critical Keywords for Big Data* (pp. 191–199). Cambridge, MA: MIT Press.

Annette Markham

26 Ethics

Abstract: Digitally saturated field sites present particular ethical challenges, often not accounted for in traditional ethics regulations or guidelines. When guidelines are contradictory or inappropriate, a core strategy is to return to core principles, seek guidance from experienced researchers, and build bespoke guidelines that favor contextual over regulatory integrity. This does not advocate dismissing extant legal and regulatory parameters, but rather, recognizes that ethics are not static, but a matter of decision making at critical junctures. A complicated situation is presented to showcase how a 'critical junctures' approach to ethics adapts to the needs of the situation, a useful consideration in criminology, since most studies will be assessed as risky, yet these tricky situations are exactly the topics that may be most important to study.

Keywords: ethics, proactive ethics, research design, ethical review, critical junctures, risky contexts

'Ethics.' This is a complicated term in academic research contexts. Not only does it conjure up vague philosophical concepts about morality and values, it also becomes a stand-in for the process of sifting through complicated regulations, taking mandatory training modules, and checking off various boxes to demonstrate one's adherence to ethical principles. These background encounters with the concept of 'ethics' are unfortunate in situations where the ethical needs are not satisfied by the ethical training one has received. When studying radical and potentially violent communities and wanting to remain in the shadows as a lurker rather than a member of the organization or a recognized researcher, how does one satisfy the regulatory norm for gaining 'informed consent?' When one seeks to learn more about successful suicide prevention by scraping confidential data from hotline calls, without asking for express permission, how can researchers justify a violation of potential privacy? How can callers in the future trust that the system is confidential? What if you want to talk with young adults online about their drug use but they can't prove they're an adult without giving away personally identifiable information? What should a researcher do if they need to study the spontaneous reactions of people to surveillance robots in public spaces, but the local ethics review committee requires informed consent in advance of exposure to the robot by default, thus negating the 'spontaneous' part of the study's design? If one wants to record and study children interacting with a controversial robot in a public square without seeking permission from them or their parents, how can one adhere to international ethics principles that tend to label children as automatically vulnerable populations?

These are complicated and actual research situations that my colleagues and I have faced that are equally relevant for those studying crime and crime control. This chapter focuses on a specific research project related to the last two questions mentioned

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-027

above, when I led a team of early career researchers in a study of public reactions to semi-autonomous robots that have been used for surveillance and control. In such a case, issues arise around key research ethics concepts, including informed consent, vulnerability, privacy protection, and data security (see Privacy and Data Protection by Bygrave).

While the case is not specific to criminology, the issues are equally relevant for those studying crime and crime control.

Readers can find many documents elsewhere to guide basic thinking about ethical regulations and ethical decision-making. Here, I focus on some basic factors one should consider when the situations are unclear, where regulatory guidelines are inadequate, or where the needs of the research situation or the local context do not match taken-for-granted norms or policies. In contexts where the company or institution's ethical guidelines seem 'off base,' or even counterintuitive, how do ethical researchers make choices?

This chapter suggests that when ethical guidelines are contradictory, not effective, or inappropriate, a core strategy is to return to core principles and seek guidance from experienced researchers. This can help researchers build bespoke guidelines that favor contextual over regulatory integrity. This does not mean that one can ignore the laws that apply in various research situations. But it recognizes that ethical guidelines are negotiable and changing, not static, and that different situations and stages of inquiry require precise and mindful application of ethical decision-making, rather than rote adherence to policies.

Ethical guidelines are negotiable and changing, not static

Developments in digital media technologies in the early 1990s transformed basic modalities of research from analog to digital forms of interaction, fieldwork, data collection, and data storage, which necessitated equivalent changes in ethical guidelines. These shifts are ongoing and challenging. Ethical guidelines for scientific study of humans were developed in response to egregious violations of human rights in biomedical research in World War II. Basic principles of ethical treatment of humans were codified in notable historical documents such as the *Nuremberg Report*, the *Helsinki Agreement*, and later, the *Belmont Report*. Concepts such as respect, justice, and beneficence were commonly specified through actions or practices that could be regulated—that is, actions that were demonstrable and measurable, such as 'obtaining informed consent,' 'protecting privacy through anonymization,' or 'excluding vulnerable categories of people from research studies.' These operationalizations were developed to ensure ethical human treatment in biomedical and psychological research contexts like intravenous drug testing or psychological experiments. These regulatory norms have been critiqued for decades by social science and humanities research communities.

They are operationalized in ways that are ill-suited for qualitative social research. For example, fieldwork studies carry different possible harms than experiments or interventions in controlled environments. Conducting research online is essentially different from face-to-face contexts and requires thinking differently about what constitutes vulnerability and consent (see Researching Online Forums by Šupa). Digitally saturated or data-implicated research contexts add many new complications to how data should be managed and protected.

Put differently, while the core principles remain broad and applicable, the operationalization of these through regulated practices and norms are narrow, specific, and often outdated. For example, the traditional practice and norm of protecting privacy of participants by anonymizing their statements is simply inadequate, since the power of data analytics is sophisticated enough to re-identify persons based on fragments of information, even just the style of speaking. To this, we can add that the most commonly adopted ethical regulations come from those nations who were the first to define the concepts in writing, establish norms through specific policies, and build training models for novice researchers. While this history is simplified here, the point is that the language and policies around what counts as ethical practice have been dominated by particular stakeholders at certain points in time, and these norms spread, not least because they have immediate utility. There is an unfortunate consequence that certain definitions are inappropriately universalized. For example, 'utilitarian' perspectives have been commonly adopted in Western countries like the UK or US, whereby harm is measured as a risk ratio in relation to benefits, and the evaluation of ethical action relates to the consequences or effects. In other words, whatever benefits the greatest number of people will outweigh the risk of harm to a few. This utilitarian or consequentialist stance clashes directly with the more 'deontological' ethical perspectives adopted in the Nordic regions, whereby, in alignment with philosophies of Immanuel Kant, one should evaluate the action in relation to the nature of the action itself. This means that sometimes, research simply should not be done.

Digital media and data analytic methods have resulted in fruitful transformations in how ethical research is defined and enacted. This is not discussed in depth here, but it is worth mentioning that at a baseline level, there is strong recognition that ethical research design does not come in a one-size-fits-all proposition. Instead, as the Association of Internet Researchers (AoIR) has argued since 2002, ethical decision-making should be grounded in the particularities of the social and technical contexts, which means paying attention to a range of considerations that continue to evolve as new technologies emerge.

When regulatory, situational, and disciplinary definitions don't match

Doing the right thing. Avoiding the creepy factor. Making ethical choices is often a matter of weighing many competing factors in specific situations with many stakeholders and interested parties. Most regulatory guidelines are idealized for non-actual situations. Thus, even if you want to maintain an ethical stance that emphasizes the importance of being sensitive to the specific contexts under study, this sort of "contextual integrity" (Nissenbaum, 2010) may be challenging. There can be large variation and contradicting rules or norms between international regulatory bodies, local or regional ethical review committees, the ethics delimited by Terms of Service in digital platforms, and disciplinary or professional ethics.

For example, in Australia or the US, two environments I've become familiar with, researchers working with people or 'human subjects' are required to submit their research design to an ethics committee and gain approval prior to carrying out the research. In many countries this would be the case when criminologists study digilantism (see Digilantism by Trottier), online violence (see Abuse by McAlinden), or sex work (see Sex Work by Rand), for example. One common expectation is that before talking with or collecting data from persons, the researcher will inform the participant of the study's purpose, ask them to confirm they comprehend, and obtain written informed consent to be a participant in the study. On the surface level, this all seems quite sensible. In actual practice, this regulatory requirement may not only be counterproductive in the field but might also be dangerous to the researchers or participants, for example when one needs to enter networks of violent behavior online in order to understand their social parameters. Informed consent from a named individual might be impossible (if you're interacting with persons in anonymous online spaces). Informed consent may be counterproductive and produce negative reactions from potential participants (if one is studying high risk activities like hacking or illegal drug trades, for example). Or, asking for consent can increase risks to the researcher (if you're studying a hate crime network and announce yourself as a researcher, you may increase the risk of becoming a victim, for example). There are many reasons why consent, or informed consent, may not be necessary or warranted. This is only one example among many, part of longstanding discussions among ethics scholars (cf., Markham, 2018, Zimmer, 2018) about the serious mismatches between different models for how ethical research can or should be accomplished.

The case: studying a controversial robot in a risky setup

Complex digital situations often involve multiple and different ethical dilemmas. Because digitally saturated contexts will most likely either occur in, or intersect with, the use of digital platforms, attention on Terms of Service is crucial, especially around issues of data scraping or announcing oneself as a researcher. Local laws around gathering, storing, or analyzing data may also be relevant, such as the EU's *General Data Protection Regulations* (GDPR), a broad set of requirements that must be attended to if collecting data from EU participants or in EU contexts. If one is conducting research at an institution like a university, there will be norms as well as policies around research integrity, data management, and research ethics review boards. These three entities are often conflated, or the committees combined, but each has special considerations. The latter, research ethics review committees, is relevant to this case since it required considerable attention.

The case I present illustrates some of these complexities. Since the study's design is too large to discuss in detail, I describe it through the potential ethical dilemmas it immediately presented to me as the research leader:

Controversial object: The research team sought to study reactions of people to a controversial robot developed by Boston Dynamics, called 'Spot.' These imposing agile and semi-autonomous robots were designed for multiple uses, "such as inspecting a bomb, rummaging through remnants of an explosion or fire, or even deescalating a potentially dangerous situation" (Bushwick, 2021). In 2021, "officers deployed the robot in just a few cases, including a hostage situation in the Bronx and an incident at a public housing building in Manhattan" (Bushwick, 2021). Within a short time of being deployed, the robots were pulled from the streets because of the negative public outcry. Among other reactions, Spot, or as the NYPD called it, 'Digidog' was denounced as frightening, creepy, and inappropriately deployed as a surveillance mechanism in poor communities (Bushwick, 2021).

Disrupting people unexpectedly in a public setting: We wanted to bring Spot to the largest public square in Melbourne, allow it to walk around in a seemingly autonomous fashion, and observe the immediate reactions of passersby.

Participants not defined in advance: The robot would encounter significant numbers of passersby, and because we didn't know who might pass by, there should be no restrictions on who could be included in the study.

Including children as participants: This would include any persons who might regularly be described as vulnerable, such as children.

Recording audio/video in public without prior permission: We also wanted to video/audio record the human/robot encounters as well as brief interviews with people just after they encountered the robot.

Data management; transferring sensitive facial image data using non-secure cloud storage: Because we wanted to observe the encounter from a variety of angles, we

would use a team of researchers, all of whom would be recording their observations through their own personal mobile phone cameras. This latter point is relevant since they would be then transferring video data of people to a central location for analysis.

Nonobvious surveillance video recorded by robot: Finally, we would also store and analyze the visual data generated by the robot's many cameras.

This research design was selected specifically to understand how people react to robots in the field, rather than in controlled experimental environments. Studying the granular moment of the human/robot interaction can yield information that is essential to build more ethical and appropriate robot designs for public spaces, especially since there are very few regulations about how robots should look or act in public spaces. This design gathers as much granular level visual and affective data as possible to help researchers and robotics designers understand how people react when they react.

The desired setup raised ethical quandaries under almost every 'red flag' category of human subject research. This includes, as a partial list: a) disrupting a person's normal expectations for walking through a public space, b) protecting participants' privacy after collecting images in an era of facial recognition, c) ensuring participants not only comprehend what's happening but also consent to being studied, d) avoiding physical and psychological harm, e) respecting people's rights to participate (or not), f) giving extra protection to vulnerable people, and g) properly handling data collected. Once you start thinking about the potential complications, it's nearly enough to stop the study before it begins.

When we pair this setup with the presence of an unexplained and unexpected robot —a robot that was deemed confronting enough to be the inspiration for a *Black Mirror* episode (Greene, 2018), depicting a techno-dystopian near future scenario—several more ethical red flags emerge that must be considered. The robot is known to cause strong reactions, since it walks on four 'legs' and can be experienced as not simply agile, but also autonomous, since the remote control is small and can be up to 50 meters distant. The robot sounds mechanical, which our participants described as militant. It also has motions reminiscent of a dog, which can be experienced as quite cute. It is sometimes experienced as uncanny, a disturbing sensation that something is not quite right. Physically, the robot is heavy and if it fell on someone, could cause injury. It has many moving parts that could crush a finger if caught in its joints.

When I explained this in detail to colleagues and the research team, they suggested modifying the study so that it would be approved by the university's ethics committee. This included removing the need for video recordings, avoiding children or other vulnerable populations, and seeking written informed consent from anyone we talked to in the field.

When I presented it as an idea, not yet a proposal, to the ethics committee, they suggested it was not as risky as I was making it out to be and I could easily frame it in less risky terms to help marshal the project through the ethics approval process. I decided to take this advice, to shrink the sheer number of red flags that the committee would need to consider.

I decided to keep all the elements of the study intact and to justify the need for this study using a utilitarian framework. This latter perspective would emphasize that there is a vital need for understanding the granular physical and psycho-social dynamics of the human/robot interaction, especially since robots are swiftly being introduced in public spaces: the potential benefit of the study far outweighs the risks of the study itself.

Three decisions are highlighted here: First, I rephrased how I was presenting information to the committee, from "this is a heavy, dangerous, and controversial robot that really should not be in the public sphere" to "This controversial robot will soon roam public spaces around the world in the near future. Since there are very few regulations in this emerging area, studying reactions of people is a critical topic."

Second, despite the committee's pressure to remove children, I insisted that since children would be the most affected by increasing numbers of robots in their lifetimes, they should be allowed to participate, react to the robot, have their opinions and reactions recorded and heard by scientists.

To sell this design to the ethics committee, I provided numerous precedents for studying youth without parental consent and extensive documentation on the rights of youth to participate in research.

Third, I held firm that consent, informed or not, should be waived for this study, since getting a natural reaction would be limited if we alerted people in advance to obtain consent for the encounter. Asking for consent afterward would also be disruptive. I was willing to compromise that if people wanted to give us their reaction, they would be informed briefly about the study and asked for verbal consent. In any confusing or tricky situations, the researchers would flag me or another senior researcher. There were plenty of information sheets to give people who wanted them, and we also encouraged people to take photos of our contact information, in case they had questions or misgivings, and if they might want to have a longer conversation about the study. These information guides and plentiful options for later contact from participants provided a measure of reassurance. They also met the standard procedures for the University's ethics committee.

This project was approved and allowed many types of interventions, data collection, and inclusion of vulnerable populations, which is remarkable since in this context, ethics committees are known to be quite strict. Notably, the three points highlighted above should not be taken as the only relevant issues or steps taken in this case, but simply illustrative of the ethical process as one of making active decisions that may go against the general flow of assessments by other members of the research team or external parties.

This is an unusual but optimistic case for embracing tricky research ideas, using strategic framing, and maintaining a strong ethic that the 'best' thing to do might sometimes be the more risky option. Even if one doesn't realize it, most research designs contain elements that can be interpreted quite differently depending on how one frames the situation or the ethical guidelines. An ethics committee might apply more or fewer restrictions than the researcher would. Ethical parameters are not pre-set or universal, even if laws and regulations are, which means there is room

for interpretation. Often this means going above and beyond what is minimally required by law. Or it can also mean establishing and defending creative practices that may seem to defy one's local regulatory norms but actually do a better job of achieving the goals of the principles on which those norms have been based.

Balancing creativity and constraint

The brief case presented above is meant to complicate the current situation of adequate ethical guidance in tricky situations, not to put off researchers, but to enable a more proactive stance that is mindful and, because it is sensitive to the needs of the situation, flexibly adaptive. This is a useful consideration in criminology, since most studies will be assessed as risky, yet these are topics that are of vital importance to study.

This discussion also offers a pathway for balancing the need to make independent decisions about what is the best move, with dependence on external ethical regulations that have been developed for good reason. Importantly, while this chapter focuses on how regulations have become outdated because of changes wrought by digital technologies, we would be remiss to always assume they are too limiting, since they can also be too loose, not providing enough guidance for especially novice researchers, or in tricky situations that might benefit from significant input.

Sometimes the tricky issue of the specific case makes it necessary to develop ethical decisions beyond and outside the boundaries of regulated norms. It can be risky to take a proactive stance rather than a more conservative stance, which helps illustrate how common regulatory guidelines rarely fit contexts of actual research enactment, and therefore can seem broken. Working beyond regulations is easier when one gets buy-in from the local ethics committee or guidance from local governing bodies and international experts.

As a summary of the key takeaway points of this chapter:

- Digitally saturated field sites or mediums for interacting with participants present particular ethical challenges that are not always accounted for in traditional ethics regulations or disciplinary guidelines.
- Concepts like 'human subject,' 'privacy,' and 'consent' require definition on a case-by-case basis, rather than universally, especially since these can be defined and experienced in many different ways in digital or data contexts.
- Digitalization creates more complications for data protection, which requires not only careful planning for data management but creative considerations for protecting privacy, beyond what may be required.
- Ethical parameters are not set in stone but ever-changing and adaptable.
- While progressive planning is advised to adapt to the needs of the specific context, this creativity should be balanced with careful attention to how situations can change.

- Ethical needs will change over the course of a study; regularly attending to the ethical situation creates strong contextual integrity.
- When situations are tricky or extant ethical guidelines don't seem to fit, researchers are advised to return to core principles and seek advice from expert and experienced researchers.

Suggested reading

By far the most comprehensive and most widely adopted guidelines for ethical decision making in digital research have been developed by the Association of Internet Researchers (AoIR) and all three distinctive sets of guidelines are recommended: 2002 (Ess, et al.), 2012 (Markham and Buchanan), and 2019 (franzke, et al.). Since most readers will need to adhere to local regulations and norms, they should also pay close attention to how local or regional guidelines match the AoIR guidelines for ethical treatment of people, norms for data management, and guidelines for research integrity. In these guidelines, readers can also find precedence and guidance for designing research that adheres to international best practice but may be in some conflict with local regulatory norms.

References

Bushwick, S. (2021). The NYPD's robot dog was a really bad idea: Here's what went wrong. *Scientific American* (May 7). https://www.scientificamerican.com/article/the-nypds-robot-dog-was-a-really-bad-idea-heres-what-went-wrong/

Ess, C., & the Association of Internet Researchers Ethics Committee. (2002). *Ethical Decision Making in Internet Research.* https://aoir.org/reports/ethics.pdf

franzke, a.s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* https://aoir.org/reports/ethics3.pdf

Greene, S. (2018). 'Black Mirror': 'Metalhead' director David Slade on the influences behind the season's most terrifying episode. *IndieWire* (January 2). https://www.indiewire.com/features/general/black-mirror-metalhead-director-david-slade-robot-dog-interview-1201912565/

Markham, A. N. (2018). Afterword: Ethics as impact—Moving from error-avoidance and concept- driven models to a future-oriented approach. *Social Media + Society,* 4(3), 1–12. http://journals.sagepub.com/doi/full/10.1177/2056305118784504

Markham, A., & Buchanan, E. (2012). *Ethical Decision-Making and Internet Research: Recommendations from the AOIR Ethics Working Committee (version 2.0).* http://www.aoir.org/reports/ethics2.pdf

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Redwood City, CA: Stanford University Press.

Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media + Society,* 4, 2056305118768300. https://doi.org/10.1177/2056305118768300

Pete Fussey

27 Facial recognition

Abstract: This chapter examines one of the most controversial law enforcement tools of recent years: facial recognition technology. The chapter first charts the growth and development of this technology—from niche curiosity to its current role as a powerful surveillance technology driven by advanced AI processing—before considering the complex human rights and social justice implications of its deployment. It concludes by highlighting how facial recognition, like other advanced digital policing tools, is exerting a profound impact on police–citizen encounters and on the nature of suspicion itself.

Keywords: surveillance, human rights, suspicion

Introduction

Facial recognition technology (FRT) has become one of the most controversial forms policing tools of the 2020s. Evangelists have described it as a 'precision crime fighting tool.' Critics have variously pointed to its overhyped effectiveness, bias across demographic groups, transformatory surveillance potential, and variety of human rights implications. Regarding the latter, members of the EU Parliament voted to ban public uses of live FRT (with notable exceptions) during 2023 and the technology has been prohibited in many parts of the US. Conversely, the UK has seen accelerated uses of FRT across public policing and private retail and leisure spaces. Other forms of FRT are less controversial and increasingly instituted into everyday life. For example, since 2017 Apple has incorporated the technology into its iPhone range. Uses of facial verification at e-borders has also grown and automated tagging of faces on social media platforms has become commonplace.

This contribution discusses the growth, impact, and meaning of FRT in the context of law enforcement. It begins by separating the complex range of technologies commonly considered together. The technological development and subsequent police adoption of FRT is then discussed. The paper then considers the wider implications of this technology, addressing the impact on policing, procedural justice, and the rights of those subjected to such surveillance.

FRT in context

'Facial recognition technology' (FRT) denotes a diverse range of tools and digital architectures. Such differences in FRT are important given their differential impact. The principle division is between 'one to one' ('*1:1*') and 'one to many' ('*1:n*') forms of image comparison. In simplest terms, *1:1* comparison involves photograph of a face

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-028

being compared with an image of itself. For example, a smartphone will unlock if the computer judges sufficient resemblance between the presented face and the photograph(s) it holds of the owner. Such applications are less controversial because they do not require a large database of potential matches and matches are largely conducted with the consent of those involved. In some senses, *1:1* FRT may be considered as a form of 'verification,' rather than 'recognition.'

1:n forms are more controversial given they require a database of potential matches to function. This is the mode most typical in law enforcement applications, with databases ('watchlists') commonly including many thousands of enrolled images. For example, the 2023 London Metropolitan Police Service FRT operation at the coronation of King Charles III used a database of around 9,000 individuals, resulting in just one single arrest. *1:n* forms of FRT are further subdivided according to different forms of use. Often viewed as most controversial is 'live facial recognition' (LFR), where cameras scan the faces of passers-by while computers match these images to a police-compiled database of individuals they are interested in. Uses of LFR are limited (mostly to the UK in Europe) and this is the form of surveillance subject to EU and US bans outlined above. The second, similar, form of *1:n* FRT is often called 'Operator Initiated Facial Recognition' (OIFR) whereby software is installed on handheld devices (typically a smartphone) and street-based police officers can photograph faces to compare with a database. More widely used is 'retrospective facial recognition' (RFR), which takes an image from an event—for example, a robbery—and compares that to a database of thousands to find a match. While considered less controversial, RFR turns any recorded public gathering into a surveillance opportunity. Moreover, RFR databases usually implicate—and therefore engage the rights of—significantly more individuals than LFR. For example, US applications of RFR rely on public databases including the driving license register. This allows police to scan against the faces of 117 million American adults (The Perpetual Lineup, 2016). In the UK, during September 2023 the Policing Minister announced an intention to make two enormous databases available for police facial recognition systems. The first is the passport photograph database. 86 % of the UK population holds a passport. The second involves 16 million images held on the Police National Database. In 2012 the High Court of England and Wales ruled that many these images were held illegally by the state. This is because the retention of photographs of those arrested but not convicted was unlawful (*RMC and FJ v Commissioner of Police for the Metropolis [2012] EWHC 1681).* These images remain on the databases used by UK policing today.

More controversial still is US tech company 'Clearview,' which has contracted its services to law enforcement agencies across the world in often opaque and unaccountable circumstances. Clearview operates by allowing faces to be scanned against a vast database of images scraped from the internet. A 2023 estimate placed Clearview's database at around four billion images (Hill, 2023). Such relationships between the number of suspects and the vast database of innocent individuals raise clear questions over proportionality, the right to privacy in public space, and fundamental protections against arbitrary state intervention (see Privacy and Data Protection by Bygrave).

The development of FRT for law enforcement

Facial recognition technology has a surprisingly long history. Gates (2011) credits the first public exhibition of the technology as a demonstration by Japanese technology company NEC (who also manufacture FRT systems currently used by police in London) at the 1970 Expo in Osaka. Law enforcement uses have since been sporadic, with short-lived deployments in Florida and London around the turn of the millennium. In a familiar tale, such uses were declared to have been an enormous success, yet the fact that such schemes were quietly dropped tells a different story. This is likely because, despite the enormous hype that regularly accompanies inaugurations of police technology (see Policing by Wilson), the systems simply failed to work to expectations, were expensive, and drained resources useful for other areas of policing. Things changed around 2015 with advancements in AI-driven computer vision image analytics brought by Convolutional Neural Network (CNN) processing. This step change in capability was further advanced by the integration of this technology into internet search engines and social media platforms that enabled reverse image searching and the automated 'tagging' of online friends.

This step change in digital image analytics translated into enhanced viability of FRT for law enforcement. Enthusiastic early adopters include the U.S. Customs and Border Protection (CBP) agency and UK policing. The former began trialing FRT to verify identifies of travelers between Atlanta and Japan before extending uses to Washington Dulles airport and, since, the processing of migrants arriving at US land borders from northern Mexico. In England, the London Metropolitan Police Service (MPS) began using NEC technology at the capital's 2016 Notting Hill Carnival, a celebration of Caribbean culture and among the largest street carnivals in the world outside Rio's Mardi Gras. Since then, the MPS have conducted an accelerating number of live facial recognition operations in London with many more conducted in the city's privatized spaces and in collaboration with commercial security operators.

Implications for public uses of facial recognition technology

FRT marks a significant shift in the capability of policing but, also, in the ways surveillance itself is conceptualized and critiqued. Proponents have sought to deemphasize the intrusiveness and impact of the technology in several ways. Among these, two justifications are particularly prominent: that FRT is similar to already accepted forms of public surveillance, such as 'CCTV' surveillance cameras, and/or that it is a "precision crime fighting tool" (Metropolitan Police Service, 2023) and, by implication, only intrudes on those 'wanted' individuals sought by the system. Both arguments are briefly analyzed in turn.

The world's first legal challenge to FRT was heard in during summer 2019 by the high court of England and Wales (*R (Bridges) v Chief Constable of South Wales Police and Others [2019] EWHC 2341 (Admin))*, brought jointly by the UK human rights organ-

ization, Liberty, and petitioner, Ed Bridges. The challenge turned on Mr. Bridges' complaint that South Wales Police illegally captured his facial biometric data on two occasions, the second at protest against an arms fair in Cardiff, Wales. The thrust of the complaint was Bridges' claim for judicial review on the assertion that police uses of FRT were not compliant with Article 8 of the European Convention on Human Rights (the right to private and family life), data protection law, and the UK's equalities legislation. The case was complex, and the court originally held many aspects of the police argument before being successfully challenged on appeal in 2020 (*R (on the application of Bridges) v Chief Constable of South Wales Police [2020 EWCA 1058]).* What is notable is that even though the original High Court ruling was excessively deferential to the police, it rejected the police argument that FRT was similar to existing surveillance camera capability. Instead, the court ruled, and thus established as legal fact, that the rights of everyone passing FRT cameras are engaged, rather than solely those individuals on watchlists. As such, FRT is different to open street surveillance cameras because, among other reasons, it engages the rights of *all* passers-by.

Furthermore, FRT technology does not passively observe like other forms of public surveillance. It captures biometric data from individuals, which is a form of data afforded special status under European data protection standards on the grounds it is of a more personal nature. Moreover, for FRT to work, it needs to both *capture* and *process* data. This distinction between data capture and processing—and recognition that the latter is more intrusive—has been recognized by the European Court of Human Rights (ECtHR) in a 2019 ruling against the UK addressing police surveillance practices during protests (*Catt v. the United Kingdom,* Judgment, ECtHR, App. No. 43514/15, 24 January 2019). A subsequent and more recent July 2023 ECtHR ruling (*Glukhin v. Russia* App no 11519/20 [ECtHR, 4 July 2023]) classified retrospective facial recognition and live facial recognition as intrusive and therefore necessitating a high level of justification for their use. Accordingly, the heightened intrusion brought by FRT has been established comprehensively as a legal fact. Such distinctions also hold implications for the ways in which FRT is conceptualized, and further underscore the difficulties of theorizing 'surveillance' as an undifferentiated set of tools and practices.

Regarding performance, it is common to hear impressive claims of FRT effectiveness. For example, Europe's heaviest users of live facial recognition, the London Metropolitan Police Service, claimed in 2023 that the technology as "a precision crime-fighting tactic" (MPS, 2023). They and have previously asserted a false positive rate of between 0.00 % and 0.13 % (MPS, 2020), thus implying the system was rarely wrong. However, closer analysis of these statistics reveals greater complexity and uncertainty over LFR performance. In circumstances where evidence is provided, most claims of positive technical performance are calculated by comparing the number of faces available for matching (e. g., the large number of passers-by) with the small number of actual matches. Therefore, if 100,000 people are estimated to pass a camera and the system incorrectly identifies 100 people, the standard evaluation methodology allows users to claim the systems are 99 % effective. The potential rights interferences of

those 100 individuals incorrectly stopped are concealed. Moreover, this approach allows system failures to be recast as successes. Using this methodology, FRT will always be presented as effective, *irrespective of how it actually performs.*

The impact of FRT

The surveillance potential of advanced facial recognition technologies has wide ranging implications for the rights of citizens, disrupts longstanding formulations of suspicion, and alters the operation of policing itself.

Facial recognition and human rights in the era of digital policing

As highlighted above, FRT has been legally recognized as an intrusive form of surveillance by both domestic and international courts. It is important to note how rights-based deliberations of FRT-related surveillance harms extend beyond notions of privacy (Article 8 of the European Convention of Human Rights (EHRC)). Given how such technologies may exert a chilling effect on individuals and groups (meaning that it inhibits them from exercising their rights; e. g., Stevens et al., 2023), FRT is implicated in the freedoms of expression (Article 10) and freedom assembly and association (Article 11) (Murray, 2023). It could be further argued that the differential performance of FRT across demographic groups, the impact on police suspicion, and the absence of effective oversight also engage the prohibition from discrimination (Article 14), right to liberty and security (Article 4), due process (Article 5), and right to effective remedy (Article 13). Recognizing the breadth of fundamental rights engaged by FRT also reveals the limitations of data protection-focused approaches for regulating this technology (such as that currently pursued by proposed UK legislation, the Data Protection and Digital Information Bill, and, prior to 2023's AI Act, a favoured approach in the EU).

Common to other forms of emerging technology, the ways in which 'necessity' is calculated are crucial to the rights implications of FRT. Typically, States commonly justify temporary limitations of many of the above rights (e. g., privacy or expression) on the basis of public security. Such rights modifications are often permissible if the three-part human rights test (an adequate legal basis, legitimacy, and whether the measure is necessary in a democratic society) is satisfied. While much misunderstood, the necessity calculation typically involves consideration of the utility of the measure and its potential harms. The more nascent a technology, the less is known about either variable.

The prohibition from discrimination

Beyond 'Article 8' rights to privacy, and deserving of its own category of discussion, is the prohibition on discrimination (Article 14 of the ECHR). Accordingly, one of the most animated areas of dispute over FRT is the issue of bias (see Bias by Oswald and Paul). Particularly acute are debates over the differential performance of algorithms in matching faces across visible ethnicities, gender, and age. One touchstone study in the debate is Buolamwini and Gebru's (2018) evaluation of related technology that evidenced (intersectional) poorer performance for females from ethnic minorities. While this research has been criticized on the grounds that Buolamwini and Gebru's (2018) work focuses on gender classification, a distinct and different technology to face recognition, the issue of ethnic, gender, and age-related bias is supported elsewhere. The most authoritative source is the US National Institute for Science and Technology (2019) evaluation of 189 commercially available face recognition algorithms that found them all to operate unequally across demographic groups. From an equality and human rights perspective this would mean someone from these populations is more likely to be incorrectly identified and thus suffer arbitrary rights infringement based on membership of a demographic group.

Subverting suspicion

Much of the complexity surrounding FRT arises because it superficially resembles other forms of more accepted surveillance. As such, any changes brought by the technology may subtle, yet are far reaching in their implications. One such subtle-yet-profound shift concerns how suspicion becomes formulated. These changes may be considered so fundamental that they challenge the basic principles and legal frameworks governing permissible state surveillance practices. Analyzing how FRT is deployed in practical operational settings reveals how the technology can impact suspicion formation in at least five ways:

i. *Subverting the sequence of surveillance.* Most legal frameworks (and all human rights tests) permitting police surveillance follow the core principle that a degree of suspicion needs to exist prior to the enactment of surveillance. This is largely how surveillance tools derive legal authorization and justify their necessity in temporarily modifying suspects' rights to privacy among others. As such, 'suspicion' needs to exist before 'surveillance.' FRT techniques, such as LFR, subvert this. All faces are scanned and deemed suspicious in the first instance, before being discounted by the computer or a human operator. In this way, FRT is much more *active* than other forms of overt surveillance.
ii. *Accountability and the origins of suspicion.* The origins of suspicion formation are complex and vary wildly in robustness and substantiation, from hunches through to criminal justice standards of evidence. Key here is that, however flawed it may be, suspicion traditionally arises from human or procedural activities that, at least

in theory, may be questioned. In the UK, a computational facial recognition match is regarded as reasonable suspicion for a street-based stop and search. Key here is the way suspicion originates outside of human discretion. Added to this is the fact that FRT processing is driven by complex AI architecture, which means the origins of such decisions are not scrutable to humans.

iii. *Technologically structured suspicion.* Relatedly, since the 1960s, sociologies of policing have revealed the complex ways suspicion is framed by myriad environmental and individual factors including organization culture and prejudice. Perhaps most notable among this literature is David Matza's (1969) concept of 'bureaucratic suspicion.' Ethnographic research on operational uses of FRT has argued that the technology brings a new component, the technological framing of bureaucratic suspicion, which can work to prime officers in reinforcing assumptions of guilt (Fussey et al., 2021).

iv. *Disrupting distinctions between covert and overt surveillance.* Many legal frameworks strive to separate covert targeted surveillance from more overt forms, with more legal restrictions and stringent authorization processes applying to the former. Users of FRT often classify the technology as overt. However, FRT relies on a hidden database of suspects. It also relies on the impossibility for individuals to know if they are enrolled onto it, regardless of how visible the cameras are. This means the system relies on covert processing to function.

v. *Individual and categorical suspicion.* That FRT performs differently across demographic categories is a scientific fact. This means the technology is more likely to screen in some demographic categories (e.g., visible ethnicity) for additional scrutiny while having a higher likelihood of incorrectly matching individuals from others (and therefore subjecting them to unnecessary police stops and arbitrary engagement of their rights). The existence of categorical suspicion—particularly on the basis of ethnicity, class, and age—is well established in the sociology of policing (e.g., Campbell, 1999). FRT adds another dimension by reinforcing such discrimination and investing it with a sense of scientific objectivity.

Conclusion

Driven by advancements in AI, current forms of FRT represent a dramatic shift in police surveillance capability. Added to this is the reliance the technology has on extracting biometric data, information categorized with a higher level of sensitivity, to function. One critical feature of debates over FRT, and over advanced AI-driven policing technologies more generally, is their deployment into conditions of legal ambiguity, and of insufficient oversight. This absence of regulatory imagination holds manifold implications.

Most obviously it means the rights of those subjected to such technologies are placed at risk of arbitrary interference. Moreover, many existing legal frameworks lack the quality and reach to offer the necessary protections needed against nascent

technologies. The different cadences of rapid technological innovations that drive such technologies, and the ponderous development of frameworks to regulate them, is a much-admired problem. One uninspired form of solutionism is to assert the primacy of humans—the 'human-in-the-loop'—as a panacea. Neither can human–technological interrelationships be separated so trivially, but human action is not shorn of context, and humans rarely have the capability or motivation to objectively scrutinize complex AI-driven processes. Human adjudication is important, but its capabilities are easily overstated. Worse, it can be offered as a synonym for digital accountability, a Potemkin façade of regulation: seemingly substantial while concealing the nothingness that lies behind. Another approach has been the establishment of AI ethics standards. Frequently normative, and often legally irrelevant, such ethical frameworks have become sufficiently numerous for surveillance users to cherry pick elements that licence and legitimate their actions. Human rights standards are exposed to sustained assault in the current era of populism. Yet they currently provide the most comprehensive means to capture and mitigate the complexity of harms and remedies brought by advanced and intrusive surveillance technologies such as FRT.

Suggested reading

Andrejevic, M., & Selwyn, N. (2022). *Facial Recognition.* New York: Wiley

Fussey, P., Davies, B., & Innes, M. (2021). "Assisted" facial recognition and the reinvention of suspicion and discretion in digital policing. *British Journal of Criminology*, 61(2), 325 – 344. https://doi.org/10.1093/bjc/azaa068

Murray, D. (2023). Police use of retrospective facial recognition technology: A step change in surveillance capability necessitating an evolution of the human rights law framework. *Modern Law Review.* DOI: 10.1111/1468 – 2230.12862

References

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1 – 15

Campbell, E. (1999). Towards a sociological theory of discretion. *International Journal of the Sociology of Law*, 27, 79 – 101.

Fussey, P., Davies, B., & Innes, M. (2021). "Assisted" facial recognition and the reinvention of suspicion and discretion in digital policing. *British Journal of Criminology*, 61(2), 325 – 344. https://doi.org/10.1093/bjc/azaa068

Gates, K. (2011) *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance.* New York: NYU Press

Hill, K. (2023) *Your Face Belongs to Us: The Secretive Startup Dismantling Your Privacy.* New York: Simon & Schuster

Matza, D. (1969) *Becoming Deviant.* Englewood Cliffs, NJ: Prentice-Hall

MPS. (2020). *Metropolitan Police Service Live Facial Recognition Trials.* https://www.met.police.uk/SysSiteAssets/media/downloads/central/services/accessing-information/facial-recognition/met-evaluation-report.pdf

MPS. (2023). *MPS LFR Policy Document.* https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/policy-documents/lfr-policy-document2.pdf

Murray, D. (2023). Police use of retrospective facial recognition technology: A step change in surveillance capability necessitating an evolution of the human rights law framework. *Modern Law Review.* DOI: 10.1111/1468 – 2230.12862

NIST. (2019). *Ongoing Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects.* Washington DC: US Department of Commerce.

Stevens, A., Fussey, P., Murray, D., Hove, K., & Saki, O. (2023) 'I started seeing shadows everywhere': The diverse chilling effects of surveillance in Zimbabwe. *Big Data & Society*, 10(1).

The Perpetual Lineup. (2016) *The Perpetual Lineup: Unregulated Police Face Recognition in America.* https://www.perpetuallineup.org/

Anthony Amicelle

28 Financial crime and surveillance

Abstract: This chapter aims at presenting and questioning the main follow-the-money techniques in digital societies for crime-control and national security purposes. It shows how national and transnational policing configurations increasingly rely on digitized financial surveillance, mainly practiced by and within leading capitalist institutions that are not proper private security companies.

Keywords: algorithm, capitalism, policing, suspicion, transnational

Introduction

"I would say, 'Follow the money, Earl, because that's where it's going to be'. Unfortunately, we did not get it following the money because the records were either nonexistent or were destroyed." This sentence is generally considered as the first ever official occurrence of the catchphrase 'follow the money.' It was pronounced by the Assistant Attorney General of the Nixon administration in June 1974, during the US Senate committee hearings on the nomination of Earl J. Silbert, the first prosecutor in the Watergate scandal. The expression was eventually popularized two years later by *All the President's Men*, the award-winning movie based on Carl Bernstein and Bob Woodward's bestseller. Today, according to international norms and criminal laws, financial flows must be—and are more and more—constantly monitored and traced to detect and prosecute a wide range of criminal activities, whilst identifying and tracking terrorists and preventing the proliferation of weapons of mass destruction. Despite differences in scale and approach, financial surveillance systems share two critical commonalities. On the one hand, financial surveillance is largely mediated by digital technologies, from algorithmic devices for suspicion to mass dataveillance schemes for collecting bulk data from commercial databases (see Surveillance by Lyon). On the other hand, the operation of highly digitized financial surveillance primarily depends on the involvement of social actors and institutions who are neither public police and intelligence services, nor proper private security companies.

Follow the money in digital societies

"In a globalized world where technology allows money to move quickly anywhere, fighting money laundering is more urgent than ever. Trillions of dollars are laundered each year. That money fuels serious crime. #FollowTheMoney." This is an excerpt of the public campaign launched by the Financial Action Task Force, namely the international policy-making body against illicit financial flows (Nance, 2018). Initiated in 1989, this global policy is now formally applied in more than 200 countries and jurisdictions,

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-029

whilst covering any kind of dirty money-related crime, from petty theft to major illegal traffickings and ruling elites illegalities, as well as terrorism and proliferation financing (Levi, 2010; Amicelle and Chaudieu, 2022). In practice, it translates into everyday financial surveillance, first and foremost to spot and report suspicious transactions for public safety and national security purposes (Harvey and Lau, 2009). But what are 'suspicious transactions,' who 'follows the money,' and how in the so-called high-tech globalized world?

Financial surveillance for crime control is conducted by corporations, starting with banks as "reluctant partners" enlisted to monitor what they are intended to circulate (Favarel-Garrigues et al., 2011; Bosma, 2022). More broadly, such a legally binding vigilance is "part of a security chain, whereby commercial data are analysed, collected, reported, shared, moved, and eventually deployed as a basis for intervention by police and prosecution. In this context, private companies – including Facebook and Twitter, airlines and banks – find themselves in the frontline of fighting terrorism and other security threats" (de Goede, 2018: 25). Policing and security configurations increasingly rely on such pillars of capitalism that are not strictly speaking private security companies, including financial institutions as well as social media platforms (Crosset and Dupont, 2022), cross-border logistics and transport businesses (Nøkleberg, 2022; Glouftsios and Leese, 2023). Unlike for-profit providers of security services and technologies of which they are themselves clients, high and low policing is neither their core business nor a direct source of capital accumulation. They do not fit the canon of the private police, referring to "the various lawful forms of organized, for-profit personnel services whose primary objectives include the control of crime, the protection of property and life, and the maintenance of order" (Joh, 2004: 55). They are originally positioned in economic and financial fields distinct from the transnational field of security at large, and the related struggles over the legitimate definition, classification, and management of internal and external security threats (Bigo, 2008).

From this perspective, both state surveillance in the name of security and corporate surveillance in the name of financial gain, especially in the age of surveillance capitalism (Zuboff, 2019), are well identified phenomena in public debate. "But the fact that the government and private establishments can participate together in 'surveillant assemblages' in the name of common objectives escapes [or at least has escaped for a long time] an analytical framework in which state and commercial purposes in collecting personal information seem incompatible" (Amicelle and Favarel-Garrigues, 2012: 117). Banks are still almost exclusively seen as organizations where we can invest or borrow money, and not as eyes and ears of the security state in the financial space, echoing public campaigns urging anyone to report suspicious activities 'if they see something' (Reeves, 2017). While they remain financial services providers on behalf of their clients, banks have also become surveillance and informant bodies on behalf of law enforcement and intelligence agencies, with up to tens of millions of financial transactions reports annually in some western countries, including several hundreds of thousands of suspicious transactions reports.

To do so, financial surveillance operates in several ways, from face-to-face surveillance to big dataveillance at a distance, with a strong emphasis on the widespread but ambivalent use of algorithmic systems (see Algorithms by Leese).

First of all, part of surveillance is still conducted 'over-the-counter,' in bank branches where clients can show up for financial operations. As part of anti-money laundering/countering the financing of terrorism compliance, banking staff, from tellers to wealth managers, must be trained to detect unusual and suspicious activities when in contact with customers (Favarel-Garrigues et al., 2008; Iafolla, 2018). In this context, surveillance is mainly focused on people's interactional behavior, in light of "potential red flags" provided by State authorities, such as "client exhibits nervous behaviour" (Amicelle, 2022). Each internal alert is then processed by banks' dedicated analysts to determine whether it must be reported to the State competent authority, namely the national financial intelligence unit (Lagerwaard, 2023; Amicelle, 2020).

Secondly, financial surveillance is also conducted on the basis of risk scoring. This leads to "stratified surveillance: differentially surveilling individuals according to their [dirty money] risk score" (Brayne, 2017: 989). Unlike face-to-face surveillance in bank branches, risk-based surveillance is conducted at a distance, without any direct interaction and observation of clients whose financial operations are placed under enhanced monitoring. The latter is carried out manually through regular reviews of high-risk clients' transactions and capital movements on the basis of recorded digital trails. The aim is still the same, i. e., to detect unusual and suspicious activities, but this time in light of transactional behaviour, with "potential red flags" such as "the transactional activity (level or volume) is inconsistent with the client's apparent financial standing, their usual pattern of activities or occupational information (e. g. student, unemployed, social assistance, etc.)" (Amicelle and Iafolla, 2018).

Over the last years, both of these forms of financial surveillance have been deemed insufficient to comply with new requirements of permanent and systematic transactions monitoring. On the one hand, surveillance over-the-counter is by definition discontinuous and partial. It only occurs during occasional interactions, which additionally become increasingly rare because of the steady decline of the number of client visits in bank branches. On the other hand, risk-based surveillance at a distance is also limited to the extent that it targets roughly 1% of the banking population, i. e., the high-risk scored clients. In this context, algorithmic devices have gradually become critical actants of financial surveillance for crime control (Amicelle and Grondin, 2021), to make a difference in three, interrelated ways.

First of all, they are deployed to monitor everyone's transactions at any time, whether or not we show up in bank branches, and regardless of our risk score. Second, they contribute to the predominance of transactional dataveillance. Indeed, transactional metrics and relations come first, and may be self-sufficient to create alerts, without any necessary reference to clients' socio-demographic and financial profile. Finally, the operation of surveillance and related creation of alerts are therefore automated, and no longer manual as compared with previous—and still coexisting—financial surveillance systems. Nevertheless, human intervention and decision-making processes do

not vanish altogether, as in most configurations of algorithmic policing and security, from police patrols to penal courts (Benbouzid, 2019; Brayne and Christin, 2021), criminal intelligence and transnational security apparatuses (Chan and Bennett Moses, 2017; Bigo and Bonelli, 2019). Algorithmic devices are designed as 'recommender systems' for agents whose responsibility is to analyze each automated alert whilst undertaking additional verification before making any final decision (Bellanova and de Goede, 2022).

In light of the growing interdisciplinary literature on algorithmic systems for policing, intelligence, and security at large, their widespread use in financial surveillance does not come as a surprise. But their relative simplicity does, by contrast with public discourses and theoretical if not speculative studies about artificial intelligence security technology. To date, the so-called big data financial surveillance programs are based on quite basic if-then commands with predefined sets of actions, far from what Daniel Neyland critically depicts as the "the algorithmic drama in current academic research" (Neyland, 2019: 81). This reflects the structural tension of crime control configurations based on the interconnection of differentiated universes of practices and rationalities, from finance to economic regulation and security intelligence. In the end, this does give rise to automated and everyday mass financial surveillance whilst remaining far removed from crime-control ambitions as it is from dystopian visions of big data, with minimal consequences if any on the existing financial order (Amicelle, 2022). In addition to the global policy against 'dirty money' that covers both money laundering and terrorist financing, the prioritization of this last issue in the context of the War on Terror has also given rise to another specific form of mass financial surveillance.

Transnational financial dataveillance in the name of counter-terrorism

"By following the money, the TFTP has allowed the U.S. and our allies to identify and locate operatives and their financiers, chart terrorist networks, and help keep money out of their hands." This is one of the current official US Treasury's promotion of their Terrorist Finance Tracking Program (TFTP), launched in October 2001 while only—and controversially—unveiled in the media in June 2006 (Gonzalez Fuster et al., 2008; Amicelle, 2011; Wesseling et al., 2012). As the most important financial data-driven surveillance programs for national security purposes, the condition for the possibility of the TFTP to exist can be summed up in a single word: traceability. According to Hermitte, "surveillance, an old reality, only becomes the modern traceability when it is carried out within an organised system, the extent of which suggests that it is a genuine project for society, pursued as much by private as by public authorities" (Hermitte, 2003: 3). Regarding transnational financial flows, such an organized system of traceability has first been implemented by the Society for Worldwide Interbank Financial Telecom-

munication—SWIFT, the self-defined backbone of global financial communication—and then by the US administration in the name of the War on Terror.

"To speak of traceability involves bringing together three elements: there must be traces and therefore a medium that makes it possible to locate them; there must be a mechanism for collecting these traces; and finally there must be a structure that allows them to be processed and analysed in order to draw conclusions. Without this type of organisation, which implies a more or less assertive voluntarism, traces do in fact exist, but not 'traceability'" (Hermitte, 2003). The SWIFT messaging platform is used by 11,000 financial institutions in more than 200 hundred countries and territories to channel up to 80 % of the digital value transfers around the world, with over 40 million messages a day, for more than 8 billion messages in 2023. In this context, financial traceability has been organized by SWIFT for commercial purposes. Each message generates digital traces that are temporarily collected and stored in SWIFT data centers, with a socio-technical structure to search and analyze them if needed. In case of unexpected problem or any specific demand from financial institutions, the cooperative is then able to find digital trails in relation to channeled transnational transactions. All three elements of traceability are thus combined for quality of the systemic messaging service.

In October 2001, this commercially organized system of traceability intersected with the US administration aim of a national security organized system of traceability, i. e., the Terrorist finance tracking program on the basis of a secondary use of SWIFT messages digital trails (Amicelle, 2013; de Goede and Wesseling, 2017). The issue of secondary use "involves data collected for one purpose being used for an unrelated purpose without people's consent" (Solove, 2007: 770). US Treasury officials have justified the access and processing of bulk data from considerable number of SWIFT messages to preemptively disrupt and incapacitate terrorist suspects. More precisely, two main narratives have been used to promote the added-value of the TFTP-related financial surveillance.

On the one hand, as other security programs with dataveillance capabilities, the TFTP has been promoted in light of interconnecting mobilities, from tracing digitized financial mobilities to finding and following physical traces of human mobilities. According to oft repeated official discourses, "For example, it is possible to locate a suspect by checking when and where the suspect closed and/or opened a new bank account in a city or country other than his or her last known place of residence. This is a clear indicator that the person may have moved. […] The terrorist finance tracking program can provide key information about the movements of suspected terrorists and the nature of their expenditures" (EU Commision, 2013: 5). In other words, on the basis of digital financial trails, the added-value would be to trace back money flows to geographically track terrorist suspects. On the other hand, it has been officially promoted to map terrorist networks. As Marieke de Goede (2012) critically points out, financial connectivity is presented as the glue or binder that holds a whole terrorist network together and as such is believed to be a major source of intelligence. This primacy given to financial relations produces suspicion by association. Having sent or received money

from a known terrorist suspect casts doubts if not outright suspicions about the individual involved.

Although the efficiency and regulation—through EU–US agreements (Bellanova and de Goede, 2022)—of such a mass financial surveillance program is still a matter of major debate in terms of privacy and fundamental rights (see Privacy and Data Protection by Bygrave), information asymmetry between sovereign entities, lack of preemptive effectiveness, and questionable added value at large, the terrorist finance tracking program has continued to exist since 2001, almost the beginning of the 21st century.

Conclusion

For the last decades, 'follow the money' has thus become the official motto of the main global policy of crime control, and the underlying logic of one of the most important transnational counter-terrorism program to date. If policing refers to "the creation of systems of surveillance coupled with the threat of sanction for deviance – either immediately or by initiating penal processes" (Reiner, 2010: 5), then what is specific to contemporary policing is the increasing significance of financial surveillance systems.

- National and transnational policing configurations increasingly rely on digitized financial surveillance.
- Financial surveillance is practiced by and within leading capitalist companies, whose economic rationale is, at first sight, relatively far from criminal law and security interventions.
- The State collection, storage, and processing of large amounts of digital data at the core of global finance functioning is used to inform national security decisions.

From this perspective, *financial surveillance: who cares?* (Amicelle and Favarel-Garrigues, 2012) should no longer be a relevant research question.

Suggested reading

Amicelle, A. (2022). Big data surveillance across fields. Algorithmic governance for policing and regulation. *Big Data & Society*, 9(2).

De Goede, M. (2012). *Speculative Security: The Politics of Pursuing Terrorist Monies.* Minneapolis: University of Minnesota Press.

King, C., Walker, C., & Gurule, J. (eds.) (2018). *The Palgrave Handbook of Criminal and Terrorism Financing Law.* Cham: Palgrave Macmillan.

References

Amicelle, A. (2011). The great (data) bank robbery: The terrorist finance tracking program & the 'SWIFT Affair', *Research Questions*, 36, 1 – 27.

Amicelle, A. (2013). The EU's paradoxical efforts at tracking the financing of terrorism. From criticism to imitation of dataveillance. *Liberty and Security Series*, 56, 1 – 19.

Amicelle, A. (2020). Right of entry. The struggle over recognition in the world of intelligence. *Political Anthropological Research on International Social Sciences*, 1(2), 243 – 272.

Amicelle, A. (2022). Big data surveillance across fields. Algorithmic governance for policing and regulation. *Big Data & Society*, 9(2).

Amicelle, A., & Chaudieu, K. (2022). The most denounced, the least punished: Ruling elites, illegalisms and anti-money laundering. *International Political Sociology*, 16(3), 1 – 22.

Amicelle, A., & Favarel-Garrigues, G. (2012). Financial surveillance: Who cares? *Journal of Cultural Economy*, 5(1), 105 – 124.

Amicelle, A., & Grondin, D. (2021). Algorithms as suspecting machines: Financial surveillance for security intelligence. In D. Lyon & D. Murakami Wood (eds.), *Big Data Surveillance and Security Intelligence: The Canadian Case* (pp. 68 – 87). Vancouver: University of British Columbia Press.

Amicelle, A., & Iafolla, V. (2018). Suspicion-in-the-making: Surveillance and denunciation in financial policing. *The British Journal of Criminology*, 58(4), 845 – 863.

Bellanova, R., & de Goede, M. (2022). The algorithmic regulation of security: An infrastructural perspective. *Regulation & Governance*, 16(1), 102 – 118.

Benbouzid, B. (2019). To predict and to manage. Predictive policing in the United States. *Big Data & Society*, 6(1).

Bigo, D. (2008). Globalized (in)security: The field and the ban-opticon. In D. Bigo and A. Tsoukala (eds.), *Terror, Insecurity and Liberty* (pp. 10 – 48). London: Routledge.

Bigo, D., & Bonelli, L. (2019). Digital data and the transnational intelligence space. In D. Bigo, E. Isin, & E. Ruppert (eds.), *Data Politics: Worlds, Subjects, Rights* (pp. 100 – 122). London: Routledge.

Bosma, E. (2022). *Banks as security actors: Countering terrorist financing at the human-technology interface.* [Thesis, fully internal, Universiteit van Amsterdam].

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977 – 1008.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction. *Social Problems*, 68(3), 608 – 624.

Chan, J., & Bennett Moses, L. (2017). Making sense of big data for security. *The British Journal of Criminology*, 57(2), 299 – 319.

Crosset, V., & Dupont, B. (2022). Cognitive assemblages: The entangled nature of algorithmic content moderation. *Big Data & Society*, 9(2).

de Goede, M. (2012). Fighting the network: A critique of the network as a security technology. *Distinktion*, 13(3), 215 – 232.

de Goede, M. (2018). The chain of security. *Review of International Studies*, 44(1), 24 – 42.

de Goede, M., & Wesseling, M. (2017). Secrecy and security in transatlantic terrorism finance tracking. *Journal of European Integration*, 39(3), 253 – 269.

European Commission. (2013). *Joint Report from the Commission and the U.S. Treasury Department regarding the Value of TFTP Provided Data.* Brussels.

Favarel-Garrigues, G., Godefroy, T., & Lascoumes, P. (2008). Sentinels in the banking industry: Private actors and the fight against money laundering in France. *The British Journal of Criminology*, 48, 1 – 19.

Favarel-Garrigues, G., Godefroy, T., & Lascoumes, P. (2011). Reluctant partners? Banks in the fight against money laundering and terrorism financing in France. *Security Dialogue*, 42(2), 179 – 196.

Gonzalez Fuster, G., De Hert, P., & Gutwirth, S. (2008). SWIFT and the vulnerability of transatlantic data transfers. *International Review of Law Computers & Technology*, 22(1 – 2), 191 – 202.

Glouftsios, G., & Leese, M. (2023). Epistemic fusion: Passenger Information Units and the making of international security. *Review of International Studies*, 49(1), 125 – 142.

Harvey, J., & Lau, S. F. (2009). Crime-money, reputation and reporting. *Crime, Law and Social Change*, 52, 57 – 72.

Hermitte, M.-A. (2003). La traçabilité des personnes et des choses. Précaution, pouvoirs et maîtrise. In P. Pedrot (ed.), *Traçabilité et responsabilité* (pp. 1 – 44). Paris: Economica.

Iafolla, V. (2018). The production of suspicion in retail banking: An examination of unusual transaction reporting. In C. King, C. Walker, & J. Gurule (eds.), *The Palgrave Handbook of Criminal and Terrorism Financing Law* (pp. 81 – 107). Cham: Palgrave Macmillan.

Joh, E. E. (2004). The paradox of private policing. *The Journal of Criminal Law and Criminology*, 95(1), 49 – 131.

Lagerwaard, P. (2023). Financial surveillance and the role of the Financial Intelligence Unit (FIU) in the Netherlands. *Journal of Money Laundering Control*, 26(7), 63 – 84.

Levi, M. (2010). Combating the financing of terrorism. A history and assessment of the control of 'threat finance'. *The British Journal of Criminology*, 50(4), 650 – 669.

Nance, M. T. (2018). The regime that FATF built: An introduction to the financial action task force. *Crime, Law and Social Change*, 69(2), 109 – 129.

Neyland, D. (2019). *The Everyday Life of an Algorithm.* New York: Palgrave Macmillan

Nøkleberg, M. (2022). Expecting the exceptional in the everyday: Policing global transportation hubs. *Security Dialogue*, 53(2), 164 – 181.

Reeves, J. (2017). *Citizen Spies. The Long Rise of America's Surveillance Society.* New York: New York University Press.

Reiner, R. (2010). *The Politics of the Police, 4th ed.* Oxford: Oxford University Press.

Solove, D. (2007). 'I've got nothing to hide' and other misunderstandings of privacy. *San Diego Law Review*, 44, 745 – 772.

Wesseling, M., De Goede, M., & Amoore, L. (2012). Data wars beyond surveillance: Opening the black box of SWIFT. *Journal of Cultural Economy*, 5(1), 49 – 66.

Zuboff, S. (2019). *The Age of Surveillance Capitalism.* New York: PublicAffairs.

David S. Wall

29 Hacking

Abstract: This chapter explores hackers and hacking. It looks at the changing role of the hacker over the years from its moralistic origins in cyberpunk social science fiction to its conceptual links with cybercrimes in the advanced computing technologies of the 21st century. The focus is mainly upon cyber-dependent (criminal) attacks upon the security of networked computers rather than cyber-enabled attacks which are primarily the domain of the fraudster. This differentiation is important if policymakers and law enforcement practitioners are to successfully identify and prosecute or manage cyber-malfeasance. It is also important because the cultural history of hacking explains it as both legal and illegal, so being able to delineate between the two levels of action is significant if law enforcement and cybersecurity agencies are to be effective in establishing trust in the integrity of computer networks and everything that they represent, including their services to users.

Keywords: hacking, cybercrime, changing role of hackers

This contribution looks at hackers, the process of hacking, and the changing role of the hacker over the years from its moralistic origins in cyberpunk social science fiction to its conceptual links with cybercrimes (Wall, 2012) in the advanced computing technologies of the 21st century (see Cybercrime by Holt and Holt). It presents the variations and developments in hacking that now focus mainly upon what are colloquially referred to as cyber-dependent (criminal) attacks upon the security of networked computers rather than cyber-enabled attacks which are mainly the domain of the fraudster (see Wall, 2024: 53–58). Such an understanding of the threats that hackers pose is important if the concept is used by policymakers and law enforcement practitioners to successfully identify and prosecute or manage cyber-malfeasance. Especially as the cultural (rather than solely legal) history of hacking means that it can be both licit and illicit, and legal and illegal, thus, being able to delineate between the two levels of action is an important focus of study. Without trust in the integrity of computer networks, everything that they represent, including their services to users becomes compromised. This contribution briefly describes the changing meaning of hacking over the years, the variation of actions and tactics that hackers use, and how they learn and reproduce their hacking skills. It then looks at the motivations and attractions for offending before concluding.

Funding note: This chapter is based upon research conducted for the EPSRC funded research project "Combatting cRiminals In The Cloud" (CRITiCal – EP/M020576/1) which contributed to Chapter 4 of Wall (2024).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-030

The changing meaning of hacking over the years

A broad reading of the literature relating to hacking reveals an etymology historically linked to developments in technology. These developments range from using an axe in medieval times to hack a way through obstacles (such as a forest), to using code in more modern times to hack through security protocols of computer programs and systems, often releasing the information within or modifying it. Hacking is therefore "an unconventional or unorthodox application of technology" (Sampson, 1959; 2005) to achieve one's goals. Sampson (2005) later expanded his 1959 definition to explain that a hacker is someone who not only "avoids the standard solution," but the hack itself becomes the basic concept to the point that the hacker becomes defined by the hack. And this captures the hacker spirit that inspired the early cyberpunk social science fiction of William Gibson, Bruce Sterling, Neal Stephenson, and Bruce Bethke (and many others); all of whom were informed by theoretical and practical developments in science and technology. These narratives subsequently helped to define the field of hacking and became a moral rallying point for the early computer hackers.

The term hacker was originally associated with broad set of moral principles as expressed in, for example, the 'Hacker's Manifesto' (Blankenship, 1986, and see also Wark, 2004). Paraphrasing Blankenship, he conceived of the hacker as a self-defined determined and curious individual who explores the new electronic world seeking knowledge that has hither to been restricted from public circulation. "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." (Blankenship, 1986).

Wark (2004) expands upon Blankenship's holistic notion of the hacker and hacking, however hacking and hackers have since become almost exclusively associated with computer misuse offending by those who commit "cybercrimes against the machine," as opposed to cybercrimes that use the machine (such as frauds) or cybercrimes in the machine (content crimes) (see Wall, 2008, 2024). Attempts during the early days of the internet to use the term 'cracking' to differentiate ethical from unethical hackers were abandoned because of the term's pejorative racial connotations in the US. So, like the term cybercrime, 'hacking' and 'hackers' are contentious, but have become firmly embedded in everyday language and therefore need to be unpacked to allow them to be understood (Wall, 2024). Not least because the concepts alone say very little about the actions or motivations of the hackers or their direction of travel, for example, whether they seek to improve security or commit crime. Even some users of social media apps who deliberately break the platform rule now present themselves as hackers, which indicate how the label has, in the spirit of its cyberpunk origins, become a rallying point for those wishing to use unorthodox applications of technology to achieve their goals (Sampson, 2005). The outcome is that the modern usage of the terms, which links hacking with criminality tends to obscure both its rather principled past and its professionalized future, for example, in Pentesting (Penetration Testing),

which seeks to firmly distance itself from hackers. The modern colloquial usage, also confuses, for example, the actions of hackers who are mainly involved with breaking into systems with the actions of, say, fraudsters who use the machine to defraud their victims.

Describing the actions of hackers

For many years hackers were lauded by media as a celebration of youth genius and the pioneering spirit of America, but they have subsequently become demonized (see Chandler, 1996: 229; Duff and Gardiner, 1996; Ross, 1990: para. 4; Sterling, 1994; Taylor, 1999; Wall, 2008, 2012, and 2024; Steinmetz, 2016). Today, attempts by commentators are often made to distinguish between 'white hat hackers' who celebrate the original ethical hacking traditions and 'black hat hackers' who are driven by more unethical and even criminal, motivations (see later). The problem is that while both descriptions appear obvious, they are hard to apply in practice and formal attempts to 'color code' hackers by their activities are even harder to apply in practice. Moreover, although there remains some regard for the public service/civil liberties ethics of the ethical hacker, their skills and beliefs have arguably become co-mingled with those of the unethical hacker, and both are now largely distrusted and regarded as a threat to the point that their activities are outlawed by most computer misuse laws. As indicated earlier, the terms hacker and hacking still say little about the direction of travel of the hackers' activities. It is therefore necessary to identify in further detail the different actions that fall under the terms.

Ethical hackers, as indicated earlier, possess a high level of ethical standards which they express through specialized knowledge combined with a belief in the ethics of freedom of access to public information. From the early days of communication technologies hackers played a crucial role in the development of the internet. They tested systems and forced code writers to achieve higher standards of quality, while also lending their skills and imagination to shape the internet. Today, the ethical hacker's role is mainly to find vulnerabilities within systems and networks which could be exploited by criminals and therefore strengthen their security. Through brute force, or Black-Box, testing they work with IT departments to identify weaknesses in system and then them to help develop policies and procedures that make them safe to operate. As such, ethical hackers have become a significant part of the cybersecurity industry, which has emerged to protect systems and are distancing themselves from unethical hackers by professionalizing their role.

Unethical hackers, in contrast to their more ethical contemporaries, maliciously and illegally search for vulnerabilities within systems which they exploit for their own gain. They are driven by a range of motivations ranging from economic or political gain, excitement, seeking revenge, or simply to show their peers how skilled they are. The main distinction between ethical and unethical hackers, however, is that the latter are prepared to break the law to achieve their goals, whereas the former are not.

Grey hat hackers are groups who occupy the middle ground or grey area between the ethical and unethical positions. They vary in terms of balance of legality/illegality of their actions and also the level of maliciousness and harm they inflict and upon whom. Although, as stated earlier, attempts to color-code hackers largely fail, these grey hat hackers are different to the above. They may, for example, use illegal methods to find vulnerabilities which they then legally sell to companies offering a 'bug bounty' for knowledge about its own or competitors' security weaknesses, or they may be tempted to illegally sell this knowledge to the highest bidder. Alternatively, they may either undertake online vigilantism (digilantism; see Digilantism by Trottier) to inflict revenge on individuals or organizations who they feel offend their values or undertake various forms of political actions (called hacktivism). These actions may be carried out as part of an online moral grouping or even directly on behalf of governments. Although in such actions they often occupy the moral high ground to justify their actions, they often seem unclear as to what represents good or bad behavior or even if they have exceeded the limits of the law.

The tactics employed by hackers to breach systems and networks

Gaining access. Hackers can gain access to systems and networks by using their programming skills and knowledge of the organization of networks. Very often ethical hackers (such as penetration testers) are invited by an organization to seek weaknesses in their own systems. Prior to beginning work, the professionals will establish their legitimate rights to access the system to avoid prosecution. Unethical hackers, in contrast, can also gain access to systems and networks by scientific means, traditionally by using their expert programming skills and knowledge of the way that networks are organized. But unethical hackers are increasingly becoming less reliant upon their personal programming skills because these have been replaced by coded routines or apps and are now being sold online on fee for a service basis. This practice of buying in cybercrime services has created a cybercrime ecosystem that facilitates different types of cybercrime. This not only helps hackers scale up their levels of attack and add a layer of obscurity to evade the gaze of law enforcement, but also allows them to focus their efforts upon the bigger cybercrime picture (Wall, 2021, 2024).

A popular method to gain unauthorized scientific access to systems (using technological methods rather than social engineering) has been, and still is, by brute force. A brute force attack is where intruders probe systems for access by entering electronically created combinations of numbers and letters which eventually form the correct access credentials. This method, however, is probably more seen in popular media representations than in practice because, on the one hand, the amount of mathematical computing needed to identify the access credentials is considerable. On the other hand, the easy availability of stolen databases containing users' default usernames

and passwords are conveniently available for purchase online for prices that reflect the quality of the data. Furthermore, entry to systems can also obtained by utilizing zero-day exploits, which are weaknesses in computing systems, mostly unpatched vulnerabilities, that are unknown to the owners or operators, and which can be exploited by offenders to attack the systems.

Another method of gaining system access credentials is to launch a man-in-the-middle (or browser-in-the-middle) attack which involves the attacker positioning a replica login page between the user and the service they are trying to access to intercept communications. Whilst the victims believe that they are communicating directly with the service provider, they are also providing their login credentials and other personal information to the hacker. This information can be sold on and used to steal money, spy on victims, sabotage their communications, or corrupt their data (see Swinhoe, 2022). Alternatively, the attacker can alter the information given to the user by the service or upload malicious code if their device is attached.

Often the objective of the cybercrime is not to gain access to systems, but to prevent others from doing so and a popular method of achieving this is to mount a Distributed denial of service attack (DDoS). DDoS attacks prevent legitimate users from gaining access to their networks and systems by bombarding their access gateways with a stream of data to overload the login capacity. DDoS is used to discomfort or distract victims whilst other actions take place, such as data theft, or installing ransomware. Ransomware attacks, like DDoS attacks, also prevent customers or clients from accessing services by encrypting business systems and stealing their essential data to prevent them from operating. Once the business victim has met the attacker's ransom demand a decryption code is provided to unlock the computer. Hackers wishing to mount browser-in-the-middle, DDoS, or Ransomware attacks can buy the services from specialist providers (also a different type of hacker, see Wall, 2021, 2024) who sell their services[1] on markets located on the deep or darkweb (the ToR router—The Onion Router) which is an area of the internet that is not searchable by conventional search engines (see Darknet by Tzanetakis).

There are also a number of alternative means by which to gain illegal access to systems that rely on more 'social' than scientific means in which the hackers focus their attentions on tricking the individual user to give them access to the system rather than using technological methods. Social engineering has long been a practical method of grooming and obtaining trust from individuals or members of organizations enough to get them to give out system information and even access credentials. Within organizations are a number of groups of workers who may be vulnerable to social engineering approaches. They are usually the lower paid and less valued in the hierarchy and who may have little emotional investment in their job and the organization they work in, yet they have access credentials. There is also a different practice of targeting more

1 Please note that these 'services' also constitute a range of cybercrimes, such as gaining access to systems and others not covered here.

senior officers of organizations to get them to grant access rights. Disgruntled insiders, for example, may abuse their legitimate system access rights to obtain information or access credentials and send them electronically outside the organizational boundaries. Alternatively, insiders may intentionally or unintentionally (through being tricked by social engineering) install spyware or back doors into a system to allow others to enter. Finally, a much less dramatic and not always illegal method is to collate the open-source data (OSINT) emitted by organizations online and their websites to create potential victims' profiles (see Wall, 2024: 73–75).

Learning and reproducing hacking skills

Participation in chat groups or forums is important in the transmission of hacking skills as they enable more experienced hackers to pass down their wisdom and experience to new members of the group. The more junior and inexperienced hackers, often called 'Script Kiddies,' 'newbies,' 'wannabes,' or 'lamers' are curious and are eager to learn hacking skills. Typically, they have low levels of competency and frequently solicit help from experienced hackers who participate in specialist chat forums. They are schooled by expert members, such as, 'Gurus' who impart their considerable knowledge and experience or 'Wizards' who have deep expert knowledge on specific areas, or 'Samurai' who have experience in carrying out practical hacking tasks:

> 'Script kiddies' tend to be inexperienced and unskilled hackers who try to infiltrate or disrupt systems by running pre-made tools. They 'vandalize websites both for the thrill of it and to increase their reputation among their peers' (Mead et al., 2005).

Lemos has referred to them as 'ankle-biters' (2000). Although denigrated for their lack of IT skills, they can nevertheless cause significant disruption to their victims' computing systems and networks. Such forum participation can also contribute to, what has been termed, the 'digital drift', from thoughts into actions and minor actions into more serious offending (see further Goldsmith and Brewer, 2015; Goldsmith and Wall, 2022).

Motivations and attractions for offending

In the 21st century, the common understanding of hacking appears to have shifted from its moral and idealistic origins towards the actions of hackers' intent upon criminal gain, especially since the emergence of the cybercrime ecosystem of skilled hacking specialist activities which facilitate modern cybercrimes. However, there is also an underlying assumption that hacking is purely for economic gain, when in fact there are a number of other motivations that can either standalone or combine to drive cyber-

crime. The need to achieve commercial advantage over competitors can be a commercial driver, as can revenge for a misdemeanor by a second party. Individual hackers may be motivated by the self-satisfaction gained from achieving a goal, or the need to obtain for peer respect for their work, or to get 'Sneaky Thrills' from the activity (Katz, 1988). Alternatively, the driver could be simply to impress potential employers and a number of hackers have written viruses that contained in the source code their employment CV. This reveals a contradiction in hacking today based upon the fact that hackers are attracted by the anonymity of the internet and the fact that hackers can distance themselves from victims, and yet they want to be known for their actions and be respected for them. This segues into politically motivated cyber-conflict and espionage hackers. Whereas hackers have traditionally wanted to be noticed for their skills, even if they did not wish to be caught, cyber-spies, who seek to conduct political or industrial espionage, in contrast do not. Their goals are to enter and exit sites discreetly, avoiding detection in order to obtain restricted information such as government or trade secrets, which might, for example, enable the possessor to gain a marketplace, or political, advantage over their competitors.

During the past decade cyber-conflict has brought a sharp realism to early apocalyptic predictions that the internet would become a hot bed of espionage and terror. Conflict has attracted hackers, and the overall number of active cyber-conflict groups has increased in as various conflicts have emerged, for example, since Russia's 'special operation' in Ukraine in 2014 and 2022 and then the Israel and Gaza conflict in 2023. The groups allied to each side in conflict comprise of sympathetic non-nation-state and nation-state actors, indicating some significant state involvement in government promoted cyber activities (Wall, 2024). These groups are very fluid and ephemeral, regularly breaking up and reforming because of internal disagreements, counter actions, deliberately evading the gaze of law enforcement. Importantly, some non-nation state (mainly criminal offenders), openly draw some credibility from geo-political affiliation in order to hide their criminality, especially economic crime. At the same time nation-state actors also exploit the economic criminals to hide their own strategic goals, not least the damage caused to the 'enemy' by cyber-campaigns and the intelligence obtained from any stolen data (see further, Wall, 2024: 69).

Four main groupings of cyber-spies exist: 'cyber-warriors' who are specialized black hat hackers acting like 'guns for hire' and selling their skills to the highest bidder; industrial spies, who focus upon enterprises and are money-driven and who infiltrate organizations and exfiltrate data; government agents, who are highly trained, skilled and seek to achieve political goals; military hackers who seek to achieve military goals.

Conclusions

Although the 'hacking' and 'hackers' were born out of moral values that emphasized curiosity, freedom of information, and curiosity for knowledge, the common understanding of the terms is now synonymous with illicit or illegal cyber-dependent activ-

ities that attack security or control systems. Not only has their meaning adapted to social and technological change over the years, but so has the variation of the actions and tactics that carry the hacking label. The latter is continuing to change via the emergence of a cybercrime service ecosystem that allows hackers to access specialist skills for a fee for which they get in return an increase in efficiency and scalability and a level of deniability to evade law enforcement. The skills-base of hacking has arguably become more complex and sophisticated as hacking skillsets have become specialized in the facilitating cybercrime ecosystem, mentioned earlier. This will arguably develop further as hackers increasingly employ generative artificial intelligence to improve the quality of their service delivery. The organization of modern cybercrime, then, reflects a business studies manual than the organized crime playbook.

Yet, the motivations for hacking and becoming a hacker have not changed and are unlikely to do so. Hackers, since the emergence of the internet, have honed their skills in chat and other online forums which also serve to reproduce and perpetuate hacking skills and maintain the hacking culture (see Accessing Online Communities by Kaufmann). This indicates that the prime motivation for hacking is not purely to commit crimes to achieve economic gain, but they vary in terms of activity, often depending upon their peer group. What is clear it that hacking has its basis in intellectual curiosity and also the pursuit of 'thrills' but can be shaped by peers into various forms of actions such as revenge, exposure (doxing), or even changing broader moral or political opinions. Finally, regardless of their legal, illicit, or illegal intentions and recalling Sampson's 1959 and 2005 commentary, hackers are characterized by their "unconventional or unorthodox application of technology." They "avoid the standard solution" and importantly, the hack itself is a function which defines the hacker.

Suggested reading

Grimes, R. (ed.) (2017). *Hacking the Hacker: Learn from the Experts Who Take Down Hackers.* New York: John Wiley & Sons, Inc.

Mitnick, K., & Simon, W. L. (2002). *The Art of Deception: Controlling the Human Element of Security.* New York: John Wiley and Sons.

Steinmetz, K. (2016). *Hacked: A Radical Approach to Hacker Culture and Crime.* New York: New York University Press.

References

Blankenship, L. (aka The Mentor). (1986). The conscience of a hacker (Hacker's Manifesto). *Phrack*, 1 (7): phile 3. Available at https://en.wikisource.org/wiki/The_Hacker_Manifesto

Chandler, A. (1996). The changing definition and image of hackers in popular discourse. *International Journal of the Sociology of Law*, 24, 229 – 251.

Duff, L., & Gardiner, S. (1996). Computer crime in the global village: Strategies for control and regulation – in defence of the hacker. *International Journal of the Sociology of Law*, 24, 211 – 228.

Goldsmith, A, Brewer, R (2015) Digital drift and the criminal interaction order, *Theoretical Criminology*, 19(1): 112 – 130.

Goldsmith, A. and Wall, D.S. (2022) The seductions of cybercrime: adolescence and the thrills of digital transgression, *European Journal of Criminology*, 19(1):98 – 117.

Katz, J. (1988). *The Seductions of Crime.* New York: Basic Books.

Lemos, R. (2000). Script kiddies: The Net's cybergangs. *ZDNet*, 13 July, www.zdnet.com/article/script-kiddies-the-nets-cybergangs.

Mead, N., Hough, E., & Stehney, T. (2005). *Security Quality Requirements Engineering (SQUARE) Methodology Report*, 31 October, Carnegie Mellon University.

Ross, A. (1990). Hacking away at the counterculture. *Postmodern Culture*, 1/1. http://jefferson.village.virginia.edu/pmc/issue.990/contents.990.html.

Sampson, P. (1959). *Dictionary of the TMRC language*, Tech Model Railroad Club of M.I.T. https://www.gricer.com/tmrc/dictionary1959.html

Sampson, P. (2005). *An Abridged Dictionary of the TMRC language*, Tech Model Railroad Club of M.I.T. https://www.gricer.com/tmrc/dictionary1959.html

Steinmetz, K. (2016). *Hacked: A Radical Approach to Hacker Culture and Crime.* New York: New York University Press

Sterling, B. (1994). *The Hacker Crackdown: Law and Disorder on the Electronic Frontier.* London: Penguin.

Swinhoe, D. (2022). Man-in-the-middle (MitM) attack definition and examples, CSO Online, 25 March, https://www.csoonline.com/article/566905/man-in-the-middle-attack-definition-and-examples.html

Taylor, P. (1999). *Hackers: Crime in the Digital Sublime.* London: Routledge.

Wall, D. S. (2008). Cybercrime and the culture of fear: Social science fiction and the production of knowledge about cybercrime. *Information Communications and Society*, 11(6), 861 – 884.

Wall, D. S. (2012). The Devil Drives a Lada: The social construction of hackers as cybercriminals. In C. Gregoriou (ed.), *Constructing Crime: Discourse and Cultural Representations of Crime and 'Deviance'* (pp. 4 – 18). London: Palgrave Macmillan.

Wall, D. S. (2021). The transnational cybercrime extortion landscape and the pandemic: Ransomware and changes in offender tactics, attack scalability and the organisation of offending, *European Law Enforcement Research Bulletin* (SCE 5), 5 October, 45 – 60.

Wall, D. S. (2024). *Cybercrime: The Transformation of Crime in the Information Age, 2nd Edition*, Cambridge: Polity.

Wark, M. (2004). *A Hacker Manifesto.* Cambridge, MA: Harvard University Press.

Robin Cameron, Gregory Stratton, and Anastasia Powell

30 Hate crime and networked hate

Abstract: The concept of hate crime is presented here as part of a wider structure of networked hate. Drawing on a digital criminological approach, this account seeks to highlight the wider structures that embolden and enable such crimes. These structures are defined by deep but less immediate and spectacular forms of harm that can be understood as a form of 'slow harm.' To avoid networked hate manifesting as hate crime we need to counter the slow violence that continues largely unaddressed in digital society.

Keywords: hate crime, violence, networks, digital, criminology

Introduction

In July 2023, the social network Threads was launched, accompanied by a proliferation of hate across its communities. Even Mark Zuckerberg, the CEO of parent company Meta, was not safe. A reply to one of his early posts contained the ominous reply: "Hey Jew." This was not a one-off post by this user, who made other racially hateful comments. Their first post on the new social media platform was: "It's the Jews," while their follow-up was a short video of notorious neo-Nazi Nick Fuentes, uttering extremely racist language. The user remains active at time of writing, despite a string of bigoted and hateful posts.

Similar patterns of hate were observable when notorious users from other platforms brought their networks of hateful infamy to Threads. White supremacist Roblox celebrity, Jon Mill, posted: "can I say the n word on this app or is that a big no no" (*sic*). His followers proceeded to spell out the racial slur, letter-by-letter, in the replies. Jack Posobiec, another white supremacist and progenitor of the 'Pizzagate' conspiracy, posted a series of hateful posts about LGBTQIA+ communities, most likely deliberate disinformation intended to test new moderation policies. Both Mill and Posobiec remain active on Twitter and Threads at time of writing, despite a string of bigoted and hateful posts.

These recent examples highlight that hateful online behavior is not the isolated actions of bigoted individuals. Rather, the rapid rise of targeted hate on Threads reflects the slow violence of accumulated networked hate, occurring through the operation and interaction of social, technical, and corporate networks. In a *digital society* various technological platforms are, on the one hand, mere tools for continuing longstanding forms of hate. They can, however, amplify, extend, and more deeply embed the harms of hate-based violence, discrimination, and exclusion; with very real, and potentially as yet unrealized consequences. In our conceptualization of Digital Criminology, *digital society* refers to an "understanding of the mutual and reciprocal shaping of technology and society" (Powell et al., 2018: 22). Digital criminology responds to this

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-031

by showing how the study of crime can move beyond 'online/offline,' 'cyber/real,' or 'virtual/terrestrial' dichotomies to recognize the "integration of the digital" in everyday life (Powell et al., 2018: 7).

In this chapter, we seek to apply a digital criminological approach to understanding both the harms of networked hate, and actions to address it. First, we elaborate the use and limitations of the conventional criminological focus on *hate crime* or *hate speech*, before supplementing the concept of *networked hate*. The normative effect of networked hate emboldens bigotry, and the cumulative harm of this hate is borne by vulnerable and marginalized people, constituting a 'slow violence.' Then, we discuss the implications of a digital criminology of networked hate that seeks to address legal, technical, and socio-cultural measures to reduce and prevent its harms (Wood, 2021). Finally, we discuss challenges that remain in the field and avenues for continued criminological research.

Limitations of 'hate crime' in a digital society

There are many terms used across criminological and socio-legal literature addressing concerns with online hate. Terms such as hate crime, cyber hate, hate speech, and hateful extremism have all been widely used for example, as well as the identification of specific issues including gender-based hate or online misogyny, cyber racism, homophobic and/or transphobic aggression, and disability-based hate. Though there is not the scope here to consider in-depth the implications of these different terms, we do find it useful to consider how criminology has traditionally engaged with online 'hate crime,' before making a case for a broadening of focus towards the harms of networked hate (Powell et al., 2018).

Though legislation varies in different countries globally, in the simplest of terms, a hate crime is a crime that is motivated by a bias against people or groups with specific characteristics that are protected by law. This often involves both a crime under a legislated crimes act, and protected characteristics under human rights or anti-discrimination legislation. Many countries have human rights and/or anti-discrimination legislation that identifies protected characteristics such as race, religion, sexual orientation, gender identity, and disability. Though the criminal element may often be a physical crime, such as homicide, assault, threats and/or vandalism, some countries have legislated specific hate crimes that are equally applicable to digitally mediated harm.

In some Australian states, for example, it is a specific criminal offense to knowingly display a Nazi symbol in public without a reasonable excuse, and there are specific laws against publicly threatening or inciting violence on grounds of race, religion, sexual orientation, gender identity or intersex or HIV/AIDS status (e.g., Crimes Act 1900 [NSW]). In this context, 'publicly' can include any form of communication such as speaking, writing, broadcasting and communicating through social media or other electronic methods to the public (see Social Media by Twigt). Additionally, many jurisdictions have specific civil law anti-discrimination protections against hate speech, which

may not constitute criminal offenses, but rather provide a mechanism for redress (such as through an apology, compensation, or other reparations).

There are, however, several limitations with hate crime, and even hate speech, as concepts for a criminological approach to the harms that emerge from the proliferation of hateful discourse in online spaces. Specifically, under these concepts the onus often remains on individuals to seek recourse when they have been specifically targeted (such as in order to report a hate crime), or to have specifically experienced a harmful incident which can form the basis of a claim for redress (such as hateful speech that resulted in lost employment, and/or psychological injury). Additionally, there are many challenges and tensions in policing and law enforcement of hate crime legislation across jurisdictions globally (Clough, 2015). For example, many vulnerable groups are not protected under current legislative frameworks (see Vulnerability by Ranchordas and Beck). Gender, for instance, may be a protected characteristic in specific anti-discrimination legislation, but not a protected characteristic under specific laws against publicly threatening or inciting violence (D'Souza et al., 2018). Additionally, protections from hate speech often come into conflict with free speech protections. Attempts to address hate at the level of individual speech acts are likely to be perceived by those it restricts as an encroachment on free speech (Powell et al., 2018, 124–126). Not only are restrictions on speech likely to further inflame such bigoted impulses, they fail to address the broader networks that sustain and embolden those views.

The 'slow violence' of networked hate

While online communications cannot be understood as the *cause* of networked hate, neither are social media and other digital platforms free of influence in co-producing and amplifying racism, misogyny, and other forms of bigotry. In other words, digital technologies are not mere tools of human action and interaction, nor are humans merely responsive to the direction of such technologies; networked hate is, like so many other practices in a digital society, inherently *technosocial.* The very human and social problem of discriminatory ideology and behaviors, combines with the very technological problem of networked communications that enable hateful content to spread with a virality and a perceived legitimacy that draws more humans in to engage with it. Research continues to show that hateful content spreads more widely and rapidly than more neutral or positive content (Daniels, 2018; Esko and Koulu, 2022). But this is no accident. The virality of networked hate is arguably driven by several key factors, such as a broader mainstreaming of extremism (Miller-Idriss, 2018), the function of algorithmic content delivery (Daniels, 2018), and the profit imperatives of technology companies themselves (Powell et al., 2018). Together these technosocial processes result in a form of 'slow violence' (Nixon, 2011; see also Brydolf-Horwitz, 2022) that, we suggest, is incrementally and insidiously causing harm. However, it does not attract the same criminological or legal recognition that arises from hate-based attacks or homicide.

Together the mainstreaming of extremism, algorithmic content delivery, and profit imperatives have had the effect of amplifying the social harms of inequality, discrimination, and hate. Yet, the severity and scale of these harms is arguably under-recognized within criminology; that is, until they erupt into criminal incidents of physical harm such as violent protests, vandalism, assaults, or fatalities. It is here that Rob Nixon's (2011) concept of slow violence could prove useful in drawing greater scholarly, and indeed public, attention to the harms of networked hate. As Nixon (2011: 2) describes: "Violence is customarily conceived as an event or action that is immediate in time, explosive and spectacular in space, and as erupting into instant sensational visibility." Slow violence, by contrast, "occurs gradually and out of sight, a violence of delayed destruction that is dispersed across time and space, an attritional violence that is typically not viewed as violence at all" (Nixon, 2011: 2). Applying this concept to the specific examples of online harms such as harassment and bullying, Brydolf-Horwitz (2022) further elaborates, highlighting the ways in which online abuse frequently fails to attract the same legal and societal outrage and action compared with physical forms of abuse.

The slow violence of networked hate has uneven impacts; and its harms are both cumulative and collective in nature. The *unevenness* of networked hate is evident in the disproportionate harms experienced by minorities and marginalized groups; such that racial, sexuality, gender identity, and disability minorities are among those most often targeted (Powell et al., 2018). These same groups already bear the brunt of discrimination in settings such as education, the workplace, healthcare, and civic participation, as well as being disproportionately targeted by physical violence and abuse (Powell et al., 2018). The *cumulativeness* of networked hate meanwhile, refers to the psychological toll that builds with each daily or regular insult, microaggression, act of exclusion, denial of personhood or identity, and that might be further interspersed with more intensive incidents such as a barrage of hateful communications, threats, blackmail, or abuse. Finally, by the *collective* nature of the harms of networked hate, we mean to highlight that marginalized and at-risk groups can still be harmed by the effects of amplified online hate and/or hate-based disinformation, even without experiencing online hate directly themselves. When hate-based discourse, misinformation, insults, and aggression towards minorities proliferate online, they reinforce and embolden these views and behaviors in other domains, such as education, workplaces, and healthcare referred to above; in effect, widening the discrimination and exclusion gap.

Preventing hate crime by addressing networked hate

As the discussion in this chapter has identified, networked hate is an inherently technosocial problem. Addressing it is, in part, hampered by the lack of urgency that comes with forms of slow violence. Among the key implications of a digital criminological ap-

proach to networked hate, is a call for research and policy leadership that recognizes the slow violence of networked hate and advocates for action to address these serious, yet not always criminal, harms. Slow violence is likely where harms arising from inequality are present. Towards this goal, integrating an awareness of, and concern for, how compounding forms of inequality perpetuate slow violence should be integrated into criminological research agendas to build greater empirical understanding how it can exacerbate the impact of crime against protected and vulnerable people. A further implication of the understanding of networked hate is that we cannot simply code our way out of this problem. In other words, given the technosocial nature of networked hate in our digital society; it will take a combination of regulatory, legal, technical, and socio-cultural solutions to address these harms.

There is arguably a role for criminologists to contribute to regulatory and legal reforms that may be needed, but also to work collaboratively outside of government agencies and with technology providers to inform technical solutions as well as community programs. Given the global scope and economic power of tech companies, national laws alone will not alone create the required change. Attempts to create transnational normative change can be seen in an initiative such as the 'Christchurch Call' in 2019, led by then Prime Minister of New Zealand Jacinda Arden as a response to the far-right terror attacks (MFAT, 2019). This led to 120 government, online service providers, and global civil society organizations signing onto a voluntary commitment to a coordinated approach to reduce violent online extremism (Christchurch Call, 2019). In the realm of regulatory and legal reform, there is a wide field of cybercrime research which acknowledges the challenges of criminal approaches to crime that occurs via global networks (Clough, 2015; Holt and Bossler, 2020). Policing resources, jurisdictional barriers, and incongruous legislation are among some of these key challenges that would apply just as readily to any legal reforms addressing networked hate.

Nonetheless, there are some promising examples, and the European Union is the clear leader in the implementation of laws and regulations that affect the behavior of large tech companies. This chapter opened with examples of networks of coordinated hate speech on Threads, however, this is not evident in the member states of the EU because at the time of writing Threads is not available there. This is because Threads is not yet compliant with regulatory laws passed by the EU, causing Meta to delay its European launch. In late-2022 the EU passed the Digital Services Act that places an obligation on online service providers to ensure that "what is illegal offline, is also illegal online" (European Council, 2021). This is the fourth in a suite of EU laws, following on from the General Data Protection Regulation, the Digital Markets Act, and the Data Governance Act intended to encourage safe, sustainable, transparent, and fair business practices by tech companies.

Meanwhile, in the Australian context the eSafety Commissioner, an independent authority, has been established by government to receive complaints of online harm. In responding to these complaints, the commissioner has been legislatively empowered to direct technology providers to remove content and impose financial penalties if take-down orders are not enacted (eSafety Commissioner, 2021). However, after writing

to Twitter or 'X' with concerns about online hate and disinformation in 2022, the Office of the eSafety Commissioner has confirmed that there are no longer any Australian-based representatives to respond to requests to remove harmful material and online hate (Evans, 2023). They now face potential fines of up to $A700,000 per day due to non-compliance with eSafety Commissioner requests (Doran, 2023). Enquiries to the press office email have been met with a 'poop emoji' reply from Twitter or 'X' (Treisman, 2023).

Technical solutions should seek to build in inclusion and safety by design. For example, *inclusion by design* refers to the intentional and proactive incorporation of diversity and accessibility considerations into the planning, development, and design of devices, services, applications, or systems (e.g., Holmes, 2020). Further, *safety by design* calls attention to identifying and mitigating potential risks and vulnerabilities during the early stages of design, rather than addressing them as an afterthought or reactively after a product or system has been developed. Together, these efforts seek to ensure that technologies are accessible, usable, and proactively address the safety needs or risks for a wide range of individuals, including those with a diversity of backgrounds. In applying these processes to the problem of networked hate, technologists could proactively consider empowering at-risk users with more control over curation and filtering of the content they receive as well as effective tools for blocking, reporting, and removal of online hate. However, it is also vital to consider mechanisms for proactively disrupting the harmful actions of users engaged in online hate. For example, through continuously improving artificial intelligence (AI) (see Artificial Intelligence by Van Brakel) algorithms to identify and remove hate speech, offensive content, and harmful behavior in real-time (Windisch et al., 2022), as well as identifying and deplatforming leaders and key influencers in hate networks (Thomas and Wahedi, 2023).

Finally, addressing networked hate also requires socio-cultural solutions aimed at tackling the root causes of hate within society through strategies that promote understanding, empathy, and inclusivity. These approaches seek to change attitudes, behaviors, and social norms to create a more equitable and harmonious society. Specific examples include fostering inclusive online communities through clear and enforced conduct standards; prevention campaigns and resources for countering hate and discrimination; supporting parents with resources for promoting digital safety and inclusion conversations at home; and digital citizenship education in schools (see, e.g., Citron and Norton, 2011).

Conclusion

Networked hate encompasses a profound but less 'immediate' and 'spectacular' form of harm than traditional hate crime. It is, we suggest, a form of slow violence that substantially shifts social norms of equality and inclusion to those of discrimination and exclusion (Nixon, 2011; see also Brydolf-Horwitz, 2022). Unfortunately, as criminologists, we have tended to focus our attention almost exclusively on explosive events,

such as when online hate transgresses into physical assault, homicide, and/or terrorism. On this basis we can say that:

- digital criminology acknowledges that the technosocial nature of our digital society requires attentiveness to the dispersed, cumulative, and often hidden forms of networked hate.
- to minimize the potential for a portion of these harms to convert to more traditionally recognizable violence, we need to counter the insidious slow violence that is growing largely unabated. Historical harms, and ongoing practices of hate-based discrimination, harassment, and violence cannot be coded away.
- meaningfully addressing the harms of networked hate will require sustained intervention into the profit-driven development of social technologies.

The seriousness of hate crimes and wider slow violence of networked hate requires a whole of society response, encompassing legal, technical, and social solutions. In addition to formal justice responses, public and private organizations need to ensure their technology use actively constructs an era of equality by design as a means of crime prevention.

Suggested reading

Brydolf-Horwitz, R. (2022). Embodied and entangled: Slow violence and harm via digital technologies. *Environment and Planning C: Politics and Space*, 40(2), 391 – 408.

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society.* London: Routledge.

Wood, M. A. (2021). Rethinking how technologies harm. *The British Journal of Criminology*, 61(3), 627 – 647.

References

Brydolf-Horwitz, R. (2022). Embodied and entangled: Slow violence and harm via digital technologies. *Environment and Planning C: Politics and Space*, 40(2), 391 – 408.

Christchurch Call. (2019). *To Eliminate Terrorist & Violent Extremist Content Online.* Christchurch Call (accessed 17 August 2023). https://www.christchurchcall.com/.

Citron, D. K., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *Boston University Law Review*, 91, 1435 – 1484.

Clough, J. (2015). *Principles of Cybercrime.* Cambridge: Cambridge University Press.

Daniels, J. (2018). The algorithmic rise of the "alt-right". *Contexts*, 17(1), 60 – 65.

Doran, M. (2023). *Australia issues Elon Musk's Twitter with a 'please explain' notice over surge in online hate.* ABC News (accessed 17 August 2023). https://www.abc.net.au/news/2023-06-22/cph-e-safety-commissioner-compels-twitter-to-produce-documents/102505546.

D'Souza, T., Griffin, L., Shackleton, N., & Walt, D. (2018). Harming women with words: The failure of Australian law to prohibit gendered hate speech. *The University of New South Wales Law Journal*, 41(3), 939 – 976.

eSafety Commissioner. (2021). *Additional Government funding to keep Australians safer online.* eSafety Commissioner (accessed 17 August 2023). https://www.esafety.gov.au/newsroom/media-releases/additional-government-funding-keep-australians-safer-online.

Esko, T., & Koulu, R. (2022). Rethinking research on social harms in an algorithmic context. *Justice, Power and Resistance*, 5(3), 307 – 313.

European Council. (2021). *What is illegal offline should be illegal online: Council agrees on position on the Digital Services Act.* European Council (accessed 17 August 2023). https://www.consilium.europa.eu/en/press/press-releases/2021/11/25/what-is-illegal-offline-should-be-illegal-online-council-agrees-on-position-on-the-digital-services-act/.

Evans, J. (2023). *Twitter has axed Australian team the eSafety regulator contacted to report child abuse material.* ABC News (accessed 17 August 2023). https://www.abc.net.au/news/2023-02-20/australian-twitter-staff-dealing-with-child-abuse-material-axed/101998040.

Holmes, K. (2020). *Mismatch: How Inclusion Shapes Design.* Cambridge, MA: MIT Press.

Holt, T. J., & Bossler, A. M. (eds.) (2020). *The Palgrave Handbook of International Cybercrime and Cyberdeviance.* London: Palgrave Macmillan.

MFAT. (2019). *Christchurch Call.* Ministry of Foreign Affairs & Trade website, accessed 17 August 2023. https://www.mfat.govt.nz/en/peace-rights-and-security/international-security/christchurch-call/.

Miller-Idriss, C. (2018). *The Extreme Gone Mainstream: Commercialization and Far Right Youth Culture in Germany.* Princeton, NJ: Princeton University Press.

Nixon, R. (2011). *Slow Violence and the Environmentalism of the Poor.* Cambridge, MA: Harvard University Press.

Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society.* London: Routledge.

Thomas, D. R., & Wahedi, L. A. (2023). Disrupting hate: The effect of deplatforming hate organizations on their online audience. *Proceedings of the National Academy of Sciences*, 120(24), e2214080120.

Treisman, R. 2023. *Got a question for Twitter's press team? The answer will be a poop emoji.* NPR (accessed 17 August 2023). https://www.npr.org/2023/03/20/1164654551/twitter-poop-emoji-elon-musk.

Windisch, S., Wiedlitzka, S., Olaghere, A., & Jenaway, E. (2022). Online interventions for reducing hate speech and cyberhate: A systematic review. *Campbell Systematic Reviews*, 18(2), e1243.

Wood, M. A. (2021). Rethinking how technologies harm. *The British Journal of Criminology,* 61(3), 627 – 647.

Malcolm Langford, Amelia Svensson, and Tone Wærstad

31 Identify theft

Abstract: Digital identity theft occupies an important place in digital criminology. The proliferation of new identification technologies and ongoing digitalization of the market and state has created a plethora of opportunities for this form of cybercrime. Scholarship on digital identity theft has grown significantly, but there is significant variation in how digital identity theft is defined, mapped, and explained, resulting in sharp divergences over how it should be tackled. In this brief entry, we conceptualize and define the phenomenon, discuss its nature, contrast competing theoretical explanations (victim-centric, offender-centric, and relational), and conclude with a discussion on policy implications and future research.

Keywords: identity theft, digital victimization, identity fraud, routine activity theory, relational cybercrime

Digital identity theft occupies an important place in digital criminology. The proliferation of new identification technologies and ongoing digitalization of the market and state has created a plethora of opportunities for this form of cybercrime (Kjørven, 2020). In 2021, it was estimated that digital identity theft cost American consumers a total of $52 billion, and burdened finance institutions with more than $200 billion in extra costs through fines, reimbursement, and security initiatives (Camplisson, 2022). In Europe, Asia, and Latin America a greater proportion of the population is affected, but the overall losses tend to be lower (Experia, 2022; ICF-SA, 2022). The advent of artificial intelligence is also raising concerns, given the reported effectiveness of machine learning-based methods in carrying out identity theft (Signicat, 2024).

Within cybercrime research, scholarship on digital identity theft has grown significantly (Bossler and Berenblum, 2019). However, there is significant variation in how digital identity theft is defined, mapped, and explained, resulting in sharp divergences over how it should be tackled. In this chapter, we conceptualize and define the phenomenon, discuss its nature, contrast competing theoretical explanations, and conclude with a discussion on policy implications and future research.

Disclaimers and acknowledgments: The chapter is written as part of the SODI project (Societal Security and Digital Identities), financed by the Research Council of Norway (project no. 320785). The authors would like to thank the editors and Marte Kjørven for comments on an earlier version.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-032

Conceptualizing identity theft

The concept of identity theft, whether digital or analog (see Digital by Wernimont), is multivalent. It has been understood and operationalized in numerous ways. In essence, the core difference is between approaches that focus on 'theft-*of*-identity' (i. e., criminal adoption, use, or trading of an identity) and 'identity-*for*-theft' (i. e., the use of identity for other crime). Following Koops and Leenes (2006), a common approach is to delineate as follows:

– *identity-related crime* (e. g., deletion or changing of an identity);
– *identity fraud* (fraud committed with identity as a target or principal tool); and
– *identity theft* (fraud or other unlawful activity in which an identity is used without consent)

The former two approaches put weight on 'theft-of-identity,' while the latter emphasizes 'identity-for-theft.'

This latter and specific approach—identity theft—is the basis of many formal definitions of digital identify theft. For example, the OECD (2009: 16) defined it as when a party without authority "acquires, transfers, possesses, or uses personal information" in connection with fraud or other crimes. Despite such clarifications, public and scholarly discourse is not always consistent, and identity theft is commonly used for all three categories (Golladay, 2020: 982).

To further complicate matters, identity theft is defined sometimes as a distinct legal offense. For instance, the US Identity Theft and Assumption Deterrence Act (title 18, s. 1028(a)(7) U.S.C.) provides that it occurs when a person who:

> [K]nowingly transfers or uses, without lawful authority, of a *means of identification* of another person with the intent to commit, or to aid or abet, *any unlawful activity* that constitutes a violation of Federal law, or that constitutes a felony under any applicable State or local law.

A similar legal definition can be found in Norway (section 202, Criminal Law Act) and an increasing number of states have created a specific offence (ICF-SA, 2022). Yet, legal clarity does not necessarily contribute to conceptual clarity. These legal definitions are both broad in their conception of identity theft (it could cover any of the above analytical categories) and narrow (by creating a specific subsidiary or residual crime of identity theft that comes in addition to the primary crime). Thus, it is important to avoid conflation of criminological and legal conceptions where they co-exist.

Finally, there is a question of what distinguishes *digital* identity theft from its *analog* cousin. Identity theft is nothing new under the sun. It simply metamorphizes with each new technological development, which has been helped by the expansion of the number of digital identities, such as electronic identity devices, usernames, logins, emails, passwords, and biometric data. The idea of digital identity theft emerged as a distinct criminological category as computers and digitalization transformed significantly the nature of the crime. It is thus most commonly invoked in situations where

the *end goal* for the misuse of identity is achieved through *digital means*, e. g., where an electronic identity device of another person is used to apply for a loan or social security benefits or gain access to a social media account to engage in criminal hate speech or blackmail (see Hate Crime and Networked Hate by Powell, Stratton, and Cameron). However, when digital identity misuse is only a *precursor* to the primary crime, it is less clear. In romance fraud, for instance, an offender will often use digital means to steal or fake an identity, but it is the victim themselves that provides, physically or digitally, the benefits, such as sending money or transferring assets (Cross and Layt, 2022). Romance fraud can be thus best understood as a digital identity-related crime.

For the purposes of this essay, we focus on the core of digital identity theft: fraud or other unlawful activity in which an identity is used digitally without consent.

Nature of identity theft

A solid body of scholarly and gray literature has sought to chart the nature of digital identity theft through representative surveys and convenience samples. Besides a focus on temporal variation (Burnes et al., 2020), this research allows cross-sectional insight into the victims of digital identity theft. While there tends to be no overall gender divide (Golladay, 2020; Brataas et al., 2022), most studies show that higher income groups are more at risk (Golladay, 2020). Although, some surveys have shown that low-income groups in high-income countries are also at risk (Williams, 2016).

Younger to middle-aged persons (between the ages of 25 and 54) tend to report more often that they have been victims. This is possibly because they are both financially and digitally active. However, there are questions over the representativity of the survey data given that children are excluded. Some scholars argue that children are the most vulnerable group as many parents inappropriately access their accounts and assets (Navarro and Higgins, 2017).

Survey data on the nature of offenders is less illuminating. Many victims are not aware of the identity of the offender or are too embarrassed or afraid to reveal it. Early US studies indicated that at least a quarter of victims knew the identity of the offender, who was a family member, acquaintance, or work colleague (Newman and McNally, 2005). In the United States, some scholars have examined reports to police and prosecutions to study offenders, and found that minority groups, younger people, and women were more commonly represented compared to physical crimes (Golladay, 2020). However, there is a risk of selection bias, as these groups (especially the former) may be (1) overrepresented in the criminal justice system due to discrimination (Brunson and Weitzer, 2009) and (2) more likely to commit 'low-tech' identity theft (e. g., stealing personal information or 'dumpster diving'—sorting through a victim's trash) —which is easier for police to detect and prosecute than 'high-tech' identity theft (e. g., hacking into websites, purchasing lists of personal information).

Turning to the effects of digital identity theft on victims, the impact is both material and non-material. The direct financial costs tend to be higher on average than

property and violent crime (Golladay, 2020: 987–988), and may be borne by a victim and/or financial institution. Resolving a digital theft incident can involve also significant time costs for all affected, which may be compounded by second-order effects. Victims report having to borrow money, change jobs/housing, and make lifestyle changes, and encountering denial of future credit, service cut-offs, and bankruptcy (ITRC, 2018). As to non-material costs, studies reveal emotional symptoms (depression, anger, anxiety), physical symptoms (headaches, trouble sleeping, high blood pressure), and challenges with trust (Golladay and Holtfreter, 2017). ITRC (2023) found that 16 % of victims considered suicide. When digital identity theft is a relational crime, it may be also accompanied by relationship strain/breakdown or threats of retaliation if the crime is reported (Langford et al., 2024).

Explaining identity theft

In seeking to explain digital identity theft, victimology literature has evolved in its understanding of explanation (see Victimization by Walklate). In our view, the theoretical landscape can be divided into three approaches: (1) victim-centric, (2) offender-centric, and (3) relational.

Victim-centric theories constitute the prevailing paradigm. Building on routine activity theory (RAT) and self-control theory (SCT), and using mostly general survey data, victim-centric approaches focus on how online behavior, personality traits, and degrees of self-guardianship correlate with risks of cybercrime victimization in a digital environment (Williams, 2016; Graham and Triplett 2017). Empirical research has examined whether targeted browsing, exposure to greater phishing through public internet use (deceiving users into revealing personal information), engagement in frequent online shopping/forums, and failure to regularly update passwords increases the likelihood of victimization.

However, the empirical findings are ambiguous (Ngo and Paternoster, 2011). Some RAT-based studies find no or weak significant relationship with certain forms of online self-guardianship, such as installing antivirus software to prevent phishing attacks (Henson, 2020). Others find that many of these guardianship activities provide some, though, limited help (see summary in Langford et al., 2024). As to SCT, Bossler and Holt (2010: 227) find that self-control had a weak relationship with password access, harassment, and having information changed, and no correlation with credit card misuse or malware infection.

A second, and smaller, body of literature is *offender-centric*. Some theories foreground how subcultural engagement encourages and shape motivations to commit cybercrime, including identity theft (Holt et al., 2017). Holt (2020: 522) argues that the "internet serves a vital role in the formation and maintenance of deviant and criminal subcultures." Relatedly, Navarro and Marcum (2020) point to social learning and the role of communication in generating deviant criminal behavior. Debutants learn from others about how to offend successfully, avoid detection, and rationalize their ac-

tions. Others point to theories grounded in cognitive and behavioral psychology, such as internet addiction (Nykodym et al., 2008; Schell, 2020) or neutralization techniques, which often involve denial of responsibility of injury, appeal to higher loyalties, and condemning the condemner (Matza and Sykes, 1957; Brewer et al., 2020).

Empirical evidence for these offender-centric theories is limited. Moreover, such theories may be more relevant to explaining cybercrimes like digital piracy and hacking (Brewer et al., 2020: 553; Holt, 2020: 520; Navarro and Marcum, 2020: 533). Digital identity theft might be simply motivated by more mundane factors. Interviews with convicted offenders, who mostly committed low-tech digital identity theft, indicated that it was driven primarily by easy financial gain, and partly 'thrill-seeking' (Golladay, 2020).

A third approach is what we call *relational cybercrime* (Langford et al., 2024). The relationship between offender and victim inflects the causes, context, and consequences of digital identity theft. Pletcher (2003: 21–22) found that offenders can take advantage of trust in close relationships in order to get a hold of sensitive security information. The intentions of offenders may vary significantly though. They might be malicious, hostile, or irrational, and based on emotional and social circumstances (Hay and Ray, 2020: 588), as would be consistent with general strain theory. Linking strain theory with feminist criminology, Marganski (2020: 636) argues that "Violating others through misogynistic vitriol, degradation, humiliation, etc., for instance, can be viewed as an act of revenge … or a means of gaining control over others so that they comply with demands." Indeed, Bossler and Holt (2010) find that when controlling for close peer victimization, the explanatory power of self-guardianship decreases suggesting that traditional victim-centric theories function poorly when an offender is known to a victim.

A relational approach also orients attention beyond the classical technological focus on phishing, hacking, and keylogging. These techniques occupy just one end of Gordon and Ford (2006)'s classification of cybercrime, so-called type I Cybercrime (which is purely digital in nature). They argued that greater focus is needed on type II Cybercrime "which has a more pronounced human element" (p. 13). This includes email and messenger services. Moreover, such type II or relational cybercrime may carry different burdens for victims. As offenders have greater access to a victim's digital devices, knowledge of their security and financial information, and greater ability to exercise physical and psychological power, they can engage more easily in "social engineering" and commit digital identity theft which has consequences that may be different or even more harmful than type I cybercrime.

Case study

The different patterns of digital identity theft, and their potential causes, can be explored through a brief case study of digital identity theft in Norway. Through our project Societal Security and Digital Identities (SODI), an analysis of the cases of 292 victims

that sought legal aid from three student legal aid clinics[1] between 2015 and 2021 was conducted (Brataas et al., 2023).

In the sample, there was a relatively even distribution of victims across the age deciles from 19 to 60; with a much lower proportion of theft for those over 60. The largest victim group was the youngest decile, with male victims tending to be younger. Overall, 64% of the victims in the sample were women, but this is partly explained by the fact that one of the clinics only serves female clients. National representative surveys tend to find no gender difference in mere incidence (NorSIS, 2022). Moreover, 68% of the digital identity theft analyzed was executed using the victim's physical electronic identity device, a single technology owned and developed by the finance industry.

In terms of effects, over half of the fraud (54%) resulted in a total loss for each victim between US$10,000 and US$500,000. This is significantly higher than the losses in most national surveys of digital identity theft (Golladay, 2020), including in Norway (NorSIS and SODI, 2024). It reflects most likely though the sample selection: victims that sought legal assistance encountered can be expected to have experienced greater harm.

Turning to the explanations and theories traversed above, there were some clear patterns. Two-thirds of victims had a close relationship with the offender: a family member, close friend, or a partner. Using an in-group analysis, we found that this pattern of victimization was gendered. As much as 80% of women had a close relationship with the offender, which was only the case for 36% of male victims. For male victims who were victims of relational cybercrime, the offender was often a work colleague. This suggests the importance of integrating feminist and gender perspectives in understanding digital identity theft.

As to technology, it varied according to the relational mode of theft. While 40% of the victims were unaware as to how their eID was misused, a third reported giving their security information to the offender *in person*, which implies significant trust. Likewise, the effects of digital identity theft also varied according to the relational modes. None of the unknown offenders stood for any fraud with a total loss over US $10,000, while 64% of the fraud carried by close peers and family resulted in a total loss of US$25,000 or more. When victims reported the case to the police, prosecution was dropped in at least 56% of the cases, with surprisingly no difference for relational cybercrime, where the identity of the offender was known. This raises questions as to why the prosecution rate is so low even when a victim can identify the offender.

Conclusion and policy implications

This entry has sought to shed light of the ever-present cybercrime of digital identity theft. Understood primarily as fraud or other unlawful activity in which an identity

1 The Law Bus (*Jussbuss*), Women's Law Clinic (JURK), and Street Lawyer (*Gatejuristen*).

is used digitally without consent, the financial and emotional costs have steadily risen during the last three decades. Criminological scholarship sought initially to identify victim behavior as a key explanation for this victimization, but the field has begun gradually to focus also on the motivations of offenders and their diverse relationships with victims.

In this respect, it is worth noting the policy implications of the prevailing victim-centric approach. RAT and SCT theories are drawn commonly on to justify delegating 'guardianship' by state and corporations to individuals (Clough, 2015; Williams, 2016: 22), with these institutions tasked with simply providing advice on how to identify and avoid risks (Hutchings and Hayes, 2009: 437). For type I 'stranger' cybercrime, this is somewhat understandable. However, it risks unnecessarily narrowing policy responses. Even worse, it may cause greater harm through victim-blaming discourses, strengthening a narrative that individuals facilitate their victimization through their lifestyle and activities (Akdemir and Lawless, 2020: 1666), which might help explain the low level of prosecutions for relational cybercrime.

Equally, victim-centric approaches risk being guided by a technological consciousness that reifies digitalization as an artefact, making it a technical challenge that can be addressed independently of systems and society in which it is embedded (Miller, 1978). Kaufmann and Jeandesboz (2017: 309) warn that, "the digital is best examined in terms of folds within existing socio-technical configurations, and as an artefact with a set of affordances that are shaped and filled with meaning by social practice." Notably, in other fields where technology is similarly dispersed and surveillance and prevention of harm is difficult, such deep social integration and the repercussions of new technologies has been acknowledged. As Yoshikawa (2018: 1157) notes, "In select fields, legislatures also augmented or replaced tort law with regulatory regimes, such as the motor vehicle safety regime, and with social insurance schemes, such as workers' compensation, to shift costs and prevent injuries." These private and public insurance-based schemes are based on the premise that the cost of technological harms should not be privatized (see Privatization by Lomell).

In our view, making a move towards a socio-technological consciousness in understanding digital identity theft is long overdue. It would avoid a singular focus on type I cybercrime and a narrow policy repertoire. More research is thus needed on the diverse forms and effects of digital identity theft, the nature of relationships between offenders and victims, and the different policies needed to address this ever-expanding cybercrime.

Suggested reading

Cross, C., Dragiewicz, M., & Richards, K. (2018). Understanding romance fraud: Insights from domestic violence research. *The British Journal of Criminology*, 58(6), 1303 – 1322.

Golladay, K. A. (2020). Identity theft: Nature, extent, and global response. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 981 – 999). Cham: Palgrave Macmillan.

Williams, M. L. (2016). Guardians upon high: An application of routine activities theory to online identity theft in Europe at the country and individual level. *British Journal of Criminology*, 56(1), 21 – 48.

References

Akdemir, N., & Lawless, C. J. (2020). Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach. *Internet Research*, 30(6), 1665 – 1687.

Bossler, A., & Berenblum, T. (2019). Introduction: New directions in cybercrime research. *Journal of Crime and Justice*, 42(5), 495 – 499.

Bossler, A., & Holt, T. (2010). The effect of self-control on victimization in the cyberworld. *Journal of Criminal Justice*, 38(3), 227 – 236.

Brataas, E. B., Svensson, A. E., & Stokke, M. (2022). Empiriske funn om misbruk av eID. In M. E. Kjørven, M. A. Hjort, & T. L. Wærstad (eds.), *Bruk og misbruk av elektronisk kommunikasjon* (pp. 115 – 129). Oslo: Karnov.

Brewer, R., Fox, S., & Miller, C. (2020). Applying the techniques of neutralization to the study of cybercrime. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 547 – 565). Cham: Palgrave Macmillan.

Brunson, R. K., & Weitzer, R. (2009). Police relations with black and white youths in different urban neighborhoods. *Urban Affairs Review*, 44(6), 858 – 885.

Burnes, D., DeLiema, M., & Langton, L. (2020). Risk and protective factors of identity theft victimization in the United States. *Preventive Medicine Reports*, 17, 101058.

Camplisson, L. (2022). Time's up: The high cost of weak digital identity. *Future Identity.* https://thefuturidentity.com/rising-fraud-weak-digital-identity/.

Clough, J. (2015). Towards a common identity? The harmonisation of identity theft laws. *Journal of Financial Crime*, 22(4), 492 – 512.

Cross, C., & Layt, R. (2022). "I suspect that the pictures are stolen": Romance fraud, identity crime, and responding to suspicions of inauthentic identities. *Social Science Computer Review*, 40(4), 955 – 973.

Experia. (2022). *2022 Global Identity and Fraud Report.* Experia.

Golladay, K. (2020). Identity theft: Nature, extent, and global response. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 981 – 999). Cham: Palgrave Macmillan.

Golladay, K., & Holtfreter, K. (2017). The consequences of identity theft victimization: An examination of emotional and physical health outcomes. *Victims & Offenders*, 12(5), 741 – 760.

Gordon, S., & Ford, R. (2006). On the definition and classification of cybercrime. *Journal in Computer Virology*, 2, 13 – 20.

Graham, R., & Triplett, R. (2017). Capable guardians in the digital environment: The role of digital literacy in reducing phishing victimization. *Deviant Behavior*, 38(12), 1371 – 1382.

Hay, C., & Ray, K. (2020). General strain theory and cybercrime. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 583 – 600). Cham: Palgrave Macmillan.

Henson, B. (2020). Routine Activities. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 469 – 489). Cham: Palgrave Macmillan.

Holt, T. (2020). Subcultural theories of crime. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 513 – 526). Cham: Palgrave Macmillan.

Holt, T. J., Freilich, J. D., & Chermak, S. M. (2017). Internet-based radicalization as enculturation to violent deviant subcultures. *Deviant Behavior*, 38(8), 855 – 869.

Hutchings, A., & Hayes, H. (2009). Routine activity theory and phishing victimisation: Who gets caught in the 'net'? *Current Issues in Criminal Justice*, 20(3), 433 – 452.

ICF-SA. (2022). *Study on Online Identity Theft and Identity-Related Crime.* Brussels: EU Commission.

ITRC. (2018). *Identity Theft: The Aftermath 2017.* El Cajon: Identity Theft Resource Center (ITRC).

ITRC. (2023). *Consumer Impact Report 2023.* San Diego: Identity Theft Resource Center.

Kaufmann, M., & Jeandesboz, J. (2017). Politics and 'the digital' From singularity to specificity. *European Journal of Social Theory*, 20(3), 309 – 328.

Kjørven, M. E. (2020). Who pays when things go wrong? Online financial fraud and consumer protection in Scandinavia and Europe. *European Business Law Review*, 31(1), 77 – 110.

Koops, B.-J., & Leenes, R. E. (2006). ID theft, ID fraud and/or ID-related crime-definitions matter. *Datenschutz und Datensicherheit*, 30(9), 553 – 556.

Langford, M., Svensson, A., & Wærstad, T. L. (2024). *Relational Cybercrime.* SODI Working Paper.

Marganski, A. J. (2020). Feminist theories in criminology and the application to cybercrimes. T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 623 – 651). Cham: Palgrave Macmillan.

Matza, D., & Sykes, G. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review*, 22(6), 664 – 670.

Miller, C. R. (1978). Technology as a form of consciousness: A study of contemporary ethos. *Communication Studies*, 29(4), 228 – 236.

Navarro, J. C., & Higgins, G. E. (2017). Familial identity theft. *American Journal of Criminal Justice*, 42, 218 – 230.

Navarro, J., & Marcum, C. (2020). Deviant instruction: The applicability of social learning theory to understanding cybercrime. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 527 – 546). Cham: Palgrave Macmillan.

Newman, G. R., & McNally, M. M. (2005). *Identity Theft Literature Review.* U.S. Department of Justice, Washington D.C.

Ngo, F. T., & Paternoster, R. (2011). Cybercrime victimization: An examination of individual and situational level factors. *International Journal of Cyber Criminology*, 5(1), 773 – 793.

NorSIS. (2022). *ID-tyveriundersøkelsen 2022: ID-tyveri rammer fortsatt mange – behov for flere tiltak og økt kunnskap.*

NorSIS and SODI. (2024). *ID-tyveriundersøkelse.*

Nykodym, N., Ariss, S., & Kurtz, K. (2008). Computer addiction and cyber crime. *Journal of Leadership, Accountability and Ethics*, 35, 55 – 59

OECD (2009), *Online Identity Theft.* Paris: OECD.

Pletcher, D. (2003). *Identity Theft: The Aftermath 2003.* San Diego: Identity Theft Resource Center.

Schell, B. (2020). Internet addiction and cybercrime. In T. Holt & A. Bossler (eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 679 – 703). Cham: Palgrave Macmillan.

Signicat (2024). *The Battle Against AI-driven Identity Fraud*, Signicat, Trondheim.

Williams, M. L. (2016). Guardians upon high: An application of routine activities theory to online identity theft in Europe at the country and individual level. *British Journal of Criminology*, 56(1), 21 – 48.

Yoshikawa, J. (2018). Sharing the costs of artificial intelligence: Universal no-fault social insurance for personal injuries. *Vanderbilt Journal of Entertainment & Technology Law*, 21, 1155.

Miria Grisot and Elena Parmiggiani

32 Infrastructures

Abstract: Infrastructures, both traditional and digital, are critical to society's functioning, with digital advancements improving efficiency but introducing challenges like cybercrime. Their study reveals the complex, interconnected sociotechnical systems formed through evolving technologies, practices, and regulations. Understanding these infrastructures is crucial for addressing contemporary digital crimes and shaping future crime control strategies.

Keywords: sociotechnical assemblage, digitalization, work practice, large-scale

Introduction

Infrastructures are critical resources for our society. Traditional infrastructures like waterways, roadways, communication lines, and electricity are foundational building blocks for today's wellbeing. In the last decades, these infrastructures have become increasingly digital, leveraging another global digital infrastructure, the internet. Simultaneously, new digital infrastructures have emerged. On the one hand, this process has arguably enabled better and more efficient services. On the other hand, it has opened new challenges in terms of protecting and securing these infrastructures from criminal threats. For example, new forms of crime such as cybercrime have emerged exploiting breakdowns in traditional (analog) and digital infrastructures (see Cybercrime by Holt and Holt).

Digital technologies have transformed enormously over the years. Born as isolated, bounded systems used in specific contexts, they have grown to encompass inter-organizational global networked 'systems of systems' which rely upon a complex set of communication standards and protocols, integration solutions, layers of hardware and software, and multiple user interfaces and use practices (Monteiro et al., 2013). The study of *infrastructures* has brought attention to how such technologies have developed over time, why and how they come into being and spread to large scale and across contexts, as well as how and by whom they are designed, made to work, cultivated, maintained, and disrupted. Studies of infrastructure have also shown what the un/intended consequences of connections making are, and they have drawn attention to the implications of the inner workings of infrastructures as knowledge generating machines (Monteiro, 2022). In addition, infrastructures are shown to be fundamentally sociotechnical and require "thinking (…) not only in terms of human versus technological components but in terms of a set of interrelated social, organizational, and technical components or systems (whether the data will be shared, systems interoperable, standards proprietary, or maintenance and redesign factored in)" (Bowker et al., 2010: 99).

In light of this, the understanding of digital technologies as infrastructures is critical for criminologists. Digital infrastructures simultaneously underpin and concern al-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-033

most all activities relevant to criminologists, ranging from financial surveillance to predictive policing, drug markets, and hacking. As a consequence, there is a need to understand digital (see Digital by Wernimont)—as well as analog—infrastructures if we want to grasp contemporary and future practices of digital crimes and crime control.

Several research fields, such as Science and Technology Studies, Information System research, Computer Supported Collaborative Work, Media Study (and also computer science, information science, communication, organization theory, cognitive science) are interested in these issues, and have developed their understanding of infrastructure. Infrastructure studies originated in the study of large sociotechnical systems, especially the work of Hughes (1987) who researched the creation/invention of the electric light and power systems/grids from the perspective of the 'infrastructure builders.' In his study, Hughes showed how infrastructure work became the site at which tensions arose, were contested and resolved. As the electricity infrastructure emerged, power shifted gradually toward those who had control of financial, material, and knowledge resources. Since the 1990s, infrastructure studies grew in importance with the advent and spread of the internet as global information infrastructure (Hanseth et al., 1996). Seminal work on infrastructures has been conducted by Star and her colleagues in the 1990s (Star and Ruhleder, 1996; Bowker and Star, 1999: 4) on research infrastructures such as infrastructures used by scientists to collaborate across research institutions. A seminal study by Star and colleagues looked at the Worm Community System, a network for geographically dispersed community of geneticists (Star and Ruhleder, 1994). One important contribution of these early studies has been to foreground the undervalued and often invisible articulation work needed to keep infrastructures reliable. In other words, as Star as pointed out infrastructure is relational, that is, the daily work of one person is the infrastructure of another (Star and Ruhleder, 1996). Over the years, infrastructure studies spread over many research fields and empirical cases from health infrastructure at hospital, regional, and national level (e.g., Ellingsen and Monteiro, 2006), to infrastructure for climate modeling (Edwards, 2013), research infrastructures, and environmental monitoring infrastructures (Parmiggiani et al., 2015).

In sum, infrastructure studies bring a unique sensibility to the understanding of digital technologies and digitalization processes and can contribute to the field of criminology by bringing a novel perspective to digital crime and crime control. In this view, students and researchers in digital criminology are offered a lens that bring attention to the opaque interdependencies across systems, people, and practices, and the resulting vulnerabilities over time and across scales (see Vulnerability by Ranchordas and Beck). In addition, infrastructure studies could inspire new investigative and research methods based on innovative data collection approaches and provide improved understanding of where, when, and why crime might occur.

Infrastructure as a lens

Infrastructure is an approach to the study of digital technologies and processes of digitalization. It is a lens to understand these phenomena as evolving sociotechnical *assemblages.* For instance, a system for electronic monitoring in prisons rather than being considered a discreet technology becomes part of a larger and evolving surveillance assemblage (Haggerty and Ericson, 2017). In this view, the multiple and different components and the practices they support come to coalesce into an assemblage with far reaching arms and effects well beyond those of each of the separate components. In this section, some pillars of infrastructure studies are introduced, followed by some methodological points on how to study infrastructure. This chapter does not have the ambition of covering all research on infrastructure, rather it points to some key concepts and key studies spanning fruitful crossovers between research fields and scholarly traditions.

Infrastructure studies favor a long-term perspective and examine how infrastructural connections are made, what shapes them and their long-term consequences. An infrastructure for sharing criminal record data for instance is complex and not built in short time. The components and connections constituting it are shaped by pre-existing decisions taken before such infrastructure is put in place. For instance, regulatory frameworks must be followed as well as technological compatibility with already installed digital tools, and existing record keeping practices. In addition, the way the infrastructure is configured will have consequence for its future development and adaptations (for instance related to novel methods for data encryption). Thus, the lens focuses on how infrastructures grow through adaptation over time and the enabling and constraining role of the existing installed base, that is, the existing and established resources, systems, practices, structures, and norms as in the criminal record infrastructure example. The concept of installed base is therefore a core construct in infrastructure studies as it points to the dependencies between past evolution and future development (Aanestad et al., 2017). Accordingly, Monteiro et al. explain infrastructure design as follows: “infrastructure design needs to serve as a link forward towards future anticipated users/uses” (2013: 580).

These processes of infrastructure development stretch beyond traditional IT projects. As an illustrative example, digital infrastructures supporting collaboration across police forces have usually been developed over time through the implementation of various information and communication technologies (see Policing by Wilson). These systems have often been procured in an uncoordinated way to address local needs of single departments and specialized tasks (such as drug enforcement, border security, or predictive policing in urban areas), and at a later point integrated into a larger supporting technological infrastructure to address issues of information fragmentation, reliability, and security (Brayne, 2020). Thus, the existing information and communication practices in the various user organizations, their structures, and regulations of professionals have shaped the technologies selected, integrated, and put into use. These

in turn will shape the future development of the infrastructure as systems are embedded into work practices. In the long term, this creates a dynamic called path dependency, that is, a tendency in which the installed base restricts the future evolution of infrastructures to specific directions, for instance by locking it with certain technological solutions and vendors (Hanseth and Lyytinen, 2010).

Infrastructures are fundamentally political. The design, development, and use of infrastructures is not neutral. Rather it is a process involving several stakeholders negotiating their divergent and often conflicting interests and aims. For instance, in a seminal study Hanseth and Monteiro (1997) have examined the politics in the making and implementation of technical standards constituting the technical backbone of an information infrastructure. Standards, they argue, are not ready made nor neutral but inscribe use behaviors. More recently Waardenburg investigated the development and implementation of new predictive policing tools in a Dutch municipality (Waardenburg et al., 2021). She found that a new profession of 'knowledge brokers' emerged with the role to translate opaque system-generated predictions into a format that was understandable and actionable for police officers. In this role, knowledge brokers acquired power in the form of more in-depth knowledge of how system-generated predictions worked and could be prepared. Similarly, Kaufmann and colleagues have shown the politics inherent in assigning form to information in digital prediction practices according to specific narratives of how crimes are committed, where to find crime, and where to send police patrols (Kaufmann et al., 2019).

In sum, these studies show how decisions that appear as technical in nature are the result of negotiations processes between different interests and between powerful and marginalized actors. Through technology design, these interests are then inscribed into durable configurations and connections that have long-lasting effects, cause tensions, and can lead to technology failures (Knobel and Bowker, 2011).

How to study infrastructures

How can infrastructures be studied given the characteristics outlined above? In other words, how does one shift focus from bonded systems (intended as artefacts, such as one computer system for predicting crime in a neighborhood) to infrastructure (intended as a complex network of systems, people, practices, norms)? In addition, how to study infrastructures which typically exist in the background, are invisible, and frequently taken for granted—while being critical to people, work, and society (Star and Ruhleder, 1994)? Also, how to tackle scale (infrastructures are typically large-scale and long-term), technical complexity and diversity (number and type of systems, their interconnections and use practices), distribution (across locations and communities of users), and not the least access? To engage in such research effort might feel overwhelming. How to find a good entry point to such complex assemblages?

In terms of scale, infrastructures have the characteristic of being simultaneously large and small—in the sense that they become visible and accessible when in contact

with daily routines and work practices. As a result, an important means to study infrastructures is to understand how they emerge in practice, that is in connection with activities and structures.

Key questions to address for researchers are: how do tensions between local and global concerns, or between individual versus community needs emerge? And how are an infrastructure's qualities distributed between the local and the global? For instance, to what extent is a metadata standard designed generic enough to represent a domain ('reach or scope") while aiming at fitting local structures, social arrangements, and technologies ('embeddedness')? A useful device to address these questions is the concept of *infrastructural inversion* (Star and Bowker, 2006) inviting researchers to shift attention from the interactions between a user and a system toward the ongoing work to sustain, upgrade, and maintain the infrastructure as a whole. For example, notable studies in the field of Science and Technology Studies propose to observe how infrastructure is navigated and made sense of by practitioners during moments of breakdown (Star and Ruhleder, 1994).

As mentioned, infrastructure is a term that invokes images of pipes, structures, and roads that interconnect and enable activities (e.g., the flow of water, electricity, cars). The understanding proposed in this chapter relies on these images while at the same time incorporating an understanding of infrastructures as also related to the flow of information and data. In the case of digital criminology, this is relevant to understand the critical role of data flow for instance to feed predictive modeling of crime occurrences. This perspective foregrounds the role of humans as designers, users, builders, owners, and in general as actors affected by infrastructural effects. Infrastructure is thus a broad term that can be defined as "to pervasive enabling resources in network form" (Bowker et al., 2010).

Conclusion

In conclusion, though infrastructure is a mature concept with origins in the study of 'old' network technologies (electricity grids, the birth of internet), its conceptual bases are still relevant. Contemporary digital technologies and processes of digitalization do not take place in a vacuum but in contexts that are already populated by multiple technologies. Thus, understanding of digital technologies as interconnecting and interconnected sociotechnical assemblages is critical.

It is important to have an understanding of digital technologies as infrastructure that is sociotechnical, and takes into consideration the organizational forms, work practices, and institutions that shape them over time, as well as their knowledge generating effects and future implications. Understanding infrastructure has implications for how we design and contribute to shaping infrastructures in the future.

Suggested reading

Parmiggiani, E. (2017). This is not a fish: On the scale and politics of infrastructure design studies. *Computer Supported Cooperative Work (CSCW)*, 26(1), 205 – 243.

Grisot, M., & Vassilakopoulou, P. (2017). Re-infrastructuring for eHealth: Dealing with turns in infrastructure development. *Computer Supported Cooperative Work (CSCW)*, 26(1 – 2), 7 – 31.

Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, 10(5), 375 – 398.

References

Aanestad, M., Grisot, M., Hanseth, O., & Vassilakopoulou, P. (eds.). (2017). *Information Infrastructures within European Health Care. Working with the Installed Base.* Cham: Springer.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (eds.), *International Handbook of Internet Research* (pp. 97 – 117). Dordrecht: Springer.

Bowker, G., & Star, S. L. (1999). *Sorting Things Out. Classification and its Consequences.* Cambridge, MA: MIT Press.

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Edwards, P. N. (2013). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.* Cambridge, MA: MIT Press.

Ellingsen, G., & Monteiro, E. (2006). Seamless integration: Standardisation across multiple local settings. *Computer Supported Cooperative Work (CSCW)*, 15, 443 – 466.

Haggerty, K. D., & Ericson, R. V. (2017). The surveillant assemblage. In D. Wilson & C. Norris (eds.), *Surveillance, Crime and Social Control* (pp. 61 – 78). London: Routledge.

Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: The case of building internet. *Journal of Information Technology*, 25, 1 – 19.

Hanseth, O., & Monteiro, E. (1997). Inscribing behaviour in information infrastructure standards. *Accounting, Management and Information Technologies*, 7(4), 183 – 211.

Hanseth, O., Monteiro, E., & Hatling, M. (1996). Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology, & Human Values*, 21(4), 407 – 426.

Hughes, T. P. (1987). The evolution of large technological systems. The social construction of technological systems. *New Directions in the Sociology and History of Technology*, 82, 51 – 82.

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.

Knobel, C., & Bowker, G. C. (2011). Values in design. *Communications of the ACM*, 54(7), 26 – 28.

Monteiro, E. (2022). *Digital Oil: Machineries of Knowing.* Cambridge, MA: MIT Press.

Monteiro, E., Pollock, N., Hanseth, O., & Williams, R. (2013). From artefacts to infrastructures. *Computer Supported Cooperative Work (CSCW)*, 22, 575 – 607.

Parmiggiani, E., Monteiro, E., & Hepsø, V. (2015). The digital coral: Infrastructuring environmental monitoring. *Computer Supported Cooperative Work (CSCW)*, 24, 423 – 460.

Star, S. L., & Bowker, G. C. (2006). How to infrastructure. In L. V. Lievrouw & S. Livingstone (eds.), *Handbook of New Media: Social Shaping and Social Consequences of ICTs* (pp. 230 – 245). London: Sage.

Star, S. L., & Ruhleder, K. (1994, October). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (pp. 253 – 264).

Star, S., & Ruhleder, K. (1996). Steps toward design an ecology and access of infrastructure: For large spaces information. *Information Systems Research*, 7(1), 111 – 134.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2021). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization Science*, 33(1), 59 – 82.

Helene Oppen Ingebrigtsen Gundhus and Jenny Maria Lundgaard

33 Intelligence

Abstract: Intelligence can be understood as an organizational activity and a power process, where security and powers are enhanced by identifying what is believed to happen in the future, in the absence of countermeasures. This chapter looks at the origins and meaning of the term and shows how intelligence can be framed both as a process, as a product, a managerial effort, and as practice. As intelligence depends on information gathering and is heavily influenced by the information technologies used, new forms of digitalization make intelligence an intriguing field for criminologists interested in the production and use of knowledge by states, in law enforcement, or elsewhere.

Keywords: datafication, intelligence, policing, security, surveillance

The term 'intelligence' is often associated with secret agents fighting enemies on behalf of a state. Whilst realities are seldom as exciting as fiction, the secretive and intrusive nature of intelligence, where information is gathered to map, reveal, predict, and prevent something unwanted, is full of important topics for criminologists to explore, especially as more of these practices are closely linked to the use and development of technological and digital platforms and systems.

Information gathering used by and against states is nothing new. With references to the Bible, ancient China, and the reign of Queen Elizabeth I, Buckley (2013) shows the constancy of such activities, as well as showcasing the problematics embedded in them, as intelligence practices can deprive citizens of their rights, instead of protecting these rights. The history and development of intelligence and intelligence agencies have been linked to the large wars of the last century (Buckley, 2013). However, current intelligence is not only a military affair. The collection of information for intelligence purposes is embedded in the activities of state departments and agencies, corporate profitmaking corporations, and non-governmental organizations (Gill and Phythian, 2018). As we will show later, intelligence practices and methods are influenced by technological advances that present us with new challenges (Clemente, 2014). As intelligence is carried out within multiple state agencies it is also directed against various phenomena, from matters of war to national security, and in the policing of various forms of crime, which is this contribution's main focus. From the early 2000s, intelligence-led policing has been implemented as a strategy in several police forces (Fyfe et al., 2018; Gundhus et al., 2022; Ratcliffe, 2016). Ratcliffe (2016) argues that this was a result of a strive towards efficiency and professionalization of policing, but geopolitical

Project funding: This research is partially funded by the Research Council of Norway, grant: 301762 and grant: 313626

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-034

factors, such as the aftermaths of the Cold War and the terrorist attacks of 9/11, have also fueled this shift, as have technological developments (Sheptycki, 2004).

Research on intelligence has been extensive within the school of international relations and defense studies, and so have historical accounts provided by former intelligence officers (Gill and Phythian, 2018). However, we find fewer historical analyses of intelligence within the field of policing (Ellefsen and Lomell, 2024). James (2013: 5) shows that intelligence activities within specialized police units in the UK date back to the 1880s. Ellefsen and Lomell (2024) use a broader concept, illustrating that intelligence practices are not only a modern concept, because systematic information gathering and reporting activities have been a police task since the first police offices were established at the end of the 17th century, showcasing how the term intelligence is rather abstract and shaped by historical connotations and origins. However, information gathering has historically often been dependent on technology. Therefore, as practices become increasingly digitalized, there is a continuum of intelligence practices and technologies, even if the digital component is new.

The elusive term 'intelligence'

In 1949 the founder of modern intelligence studies, Sherman Kent, defined intelligence as consisting of three overarching elements: intelligence is *knowledge*, produced in a *particular kind of organization or enterprise* that *performs particular intelligence activities* (Kent, 1966). We here include a fourth and often underlined element, namely that *the activities have a particular purpose or function* in mind. In the thirty definitions of intelligence found by Rønn (2016), these elements were present, but the definitions also aimed to include new perspectives. Following the traces of the concept intelligence, it stands out as dynamic, elusive, and easily adapted to various environments. There is no agreement on how to define intelligence across different contexts, but common through history seems to be "know your enemy," and something that is "a means to an end" (Gill and Phythian, 2018: 1). The most comprehensive definition and perhaps the most cited, Gill and Phythian's, is of particular interest for digital criminology. Their core concept to understand intelligence is surveillance (see Surveillance by Lyon), underscoring its connection to knowledge and power, defining it as "mainly secret activities – targeting, collection, analysis, dissemination and action – intended to enhance security and/or maintain power relative to competitors by forewarning threats and opportunities" (Gill and Phythian, 2018: 5). Though secrecy often characterizes intelligence activities and operations, the term 'mainly secret' also includes open and unclassified intelligence (Rønn, 2016). Intelligence can thus be understood as an organizational activity and a power process, where policies and actions enhance security and power relations by identifying what is believed soon to happen in the absence of countermeasures. As the availability of information is key to intelligence practices, digitalization creates a notable impact: it enables the collection, storage, and analysis of vast amounts of data. The possibility for precautionary types of intelligence has there-

fore increased tremendously, due to digitization of incidents, the accessibility of digital data, and big data analytics (Big Data by Završnik).

Framing intelligence

Within the scholarly tradition of international relations, intelligence is understood as mainly related to politics. However, particularly for criminology, intelligence is also about the making of 'crime intelligence' as part of police decision-making processes, touching upon tensions between knowledge, scientization, and politicization. Liberal democracies strive for a balance between intelligence measures and democratic values (Hillebrand, 2014: 305). Normative and ethical questions are paramount to research on intelligence, as its aims and measures are often unregulated by law (Rønn, 2016). As intelligence products lay the basis of intrusive interventions, ethical questions of trust, legitimacy, and democratic control are also crucial questions for researchers and practitioners (Diderichsen, 2016). Since the core element of intelligence is monitoring and surveillance, oversight bodies are important to ensure legitimacy.

The digital aspect of intelligence is central in the practices of information gathering and processing. Ratcliffe (2016 states that technology and intelligence-led policing are closely related, and Sheptycki shows how new information and communication technologies (ICT) fueled "the rise of 'intelligence-led policing'" (Sheptycki, 2004: 307). The information revolution of the 1990s reconfigured the police sector to take advantage of ICT, bringing intelligence and policing closer together. This 're-tooling' culminated in the introduction of various models of intelligence-led policing to improve the police's ability to collect, analyze, and disseminate crime intelligence. As intelligence practices require collection and use of data, understanding the role of digital technologies, such as sources, storage, and tools for analysis, is crucial to understand intelligence today. Before returning to these digital aspects of intelligence production in more detail, we present four ways of framing intelligence: As *a process*, *a product*, *a managerial effort*, and as *practice.* As we will show, all these approaches actualize intelligence in relation to digitalization.

The intelligence cycle is the most traditional model to conceptualize the *intelligence process* (Gill and Phythian, 2018). It presents a structured process, which in its simplest form comprise five stages: 1) planning and direction, 2) collecting, 3) processing, 4), analyzing, and 5) dissemination. According to Warner (2013), this structure came from French revolutionaries, who needed a means to explain intelligence collection to their increasingly specialized military staff. He asserts that the graphic depiction of a cycle became popular, even as governments moved beyond such a simple and linear process (Warner, 2013). The accuracy and relevance of this model has been critiqued for its simplification but is still commonly used. In line with the intelligence cycle, intelligence is often presented as the result of finding and turning data into information, which becomes intelligence when analyzed (Ratcliffe, 2016). This evolution from data to information has been challenged (Kaufmann, 2023), since data, too,

is information. Thus, data are treated in this contribution as a type of information. Bits of data are collected, and registered into various computer systems, quantified, processed, and analyzed using software before being disseminated. This datafication of intelligence (Chan et al., 2022) points to how data encoded in the intelligence system must be accounted for. With digitization, data is rendered digital and, hence, traceable. This datafication implies a process where "human actions are turned into data, often for strategic, optimization or governance purposes" (Flyverbom, 2022: 4). During such processes data is further constructed, translated, curated, and co-produced in dynamic interaction between humans and technologies, such as computer software and apps (Kaufmann et al., 2018; Lundgaard and Gundhus, 2024). This brings us to intelligence as a product.

The intelligence product is a means to an end. Its objective is to provide support for decisions on resource allocations and countermeasures. Ratcliffe (2016) defines intelligence as actionable knowledge for police managers. It is in this context the future oriented and proactive aspect of intelligence is rendered significant, underscored by terms such as 'probability' or 'likelihood.' The implementation of intelligence has been presented in terms of its scientific approach to crime analysis, though these claims of objectivity have been criticized (Vestby, 2018). Innes et al. (2005: 39) argue that the products should be recognized "as an artefact of the data and methods used in their construction, rather than providing an accurate representation of any crime problems," stating that the best intelligence products can do is to help reduce uncertainty. However, when intelligence products are disseminated, uncertainty or ambiguity is often less visible, as is any lack of essential data. Thus, the intelligence product, the basis for decision-making, may appear more certain than it is. Digitalization is also an aspect of this, as Bennett Moses and Chan (2018: 818) argue: "there is a potential accountability gap whenever the software itself becomes an acceptable basis for decision-making so that those to whom account is given do not (or are unable to) dig deeper." Intelligence, then, is the co-product of both humans and technologies, and decision-making ends up being partially outsourced to the software itself (Kaufmann, 2018).

Intelligence is also a managerial effort, and a way of governing knowledge production, as the product should be actionable and have a purpose. The implementation of intelligence-led policing was related to demands for cost efficiency, as well as for an increased focus on risk, control cultures, and securitization of social problems in the penal state (Garland, 2001). To apply scientific measures to manage and control crime was connected not only to a more scientific rationale (Innes et al., 2005), but also to managerialism and new public management. Risk-oriented institutional logics and risk assessments were also implemented in the private sector, and intelligence-led policing emerged in the 1990s at the intersection of science discourses, rationalization, and managerialism. Ericson and Haggerty (1997), as well as Innes and Sheptycki (2004) explored these new criminological fields as an issue of science and technology studies, highlighting the importance of technologies at this intersection, and arguing that such policing applies scientific concepts as a legitimizing aura of objectivity (see also Innes et al, 2005).

Intelligence as practice has been less researched than the aims, ideals, and knowledge products from intelligence. Thus, there is less research on how intelligence practices actually identify what may happen in the absence of any countermeasures. There is research on failures (see Gill and Phythian, 2018), but fewer studies on the ways in which intelligence-led policing is played out (Ratcliffe, 2016). What do the agencies do, and how do they do it? How are their practices enacted, and how do they influence intelligence at large? When looking into the practices of intelligence, it also becomes clear that there is a lot more happening than just core activities, such as collection, analysis, counterintelligence, and covert action. Much research "neglect the variety of activities that intelligence services have also engaged in, such as conducting diplomacy, guarding borders, running prisons, operating military units, designing atomic bombs, and managing professional soccer teams" (Stout and Warner, 2018: 517). Intelligence practices also have a digital dimension, as the digital systems used by intelligence agents and agencies have various affordances (Adams and Thompson, 2016), which shape and define the interactions between the technologies and the humans using them (boyd, 2011), and thus also the outcome of these interactions. The dependencies of technologies have been explored in implementation studies in the UK (James, 2013), Australia, Canada, New Zealand (Burcher and Whelan, 2019; Sanders et al., 2015; Weston, Bennett-Moses, and Sanders, 2019), Norway (Gundhus, 2013; Gundhus et al., 2022), and the US (Carter, 2013). Practices are not only reflecting human activities, but also those of the computational systems used as sources and for processing data and making intelligence products. Data systems can for example feature feedback loops, resulting in proactive policing that effects conceptions of crime conducted by people in the communities (O'Neil, 2016).

Control of political behavior to protect the state, what Brodeur (1983) calls *high policing*, relies on the storage of intelligence data. Since *low policing*, that is "routine law enforcement and street level order maintenance" (Reiner, 1985: 2), also increasingly includes intelligence methods, the need for digitization, data collection, and surveillance in day-to-day ordinary police work increases. The dilemmas posed by digitization thus mean that the increased use of high policing methods on low policing targets can impact trust and legitimacy due to a lack of accountability and oversight mechanisms (Bowling et al., 2019: 24; Brodeur, 2010).

Intelligence and digitalization

The use of intelligence in criminal justice systems speaks to broader *trends in criminal policies and practices*, such as pre-crime, prediction, and preemption, in combination with data-driven policing using various digital tools and software (McCulloch and Wilson, 2016). Data-driven approaches to information reinforce a larger trend towards the "management of visibilities," where only coded data can be used and considered a valid knowledge base (Flyverbom, 2022).

Digitalization makes intelligence an intriguing field for criminologists interested in the production and use of knowledge by states, in law enforcement, or elsewhere. Digitalization enables an increased sharing of information and is implemented in view of changing the working practices of the organizations (Chan et al., 2022). Ratcliffe claims such change is a necessary step in policing, as knowledge "must be structured in a way that can help decision-makers develop policy" (Ratcliffe, 2016: 74). Due to the explosion of openly available information (so-called OSINT), much intelligence is now taking place online, where the collection of data is often seen as low in risk and high in reward (Clemente, 2014). The availability of recorded data enables new strategies for knowledge production, while established knowledge production changes through access to new software for data integration.

Techniques for analysis change due to new software, machine learning, and language models. Such developments in software products support the move towards more surveillant policing (Brayne, 2020) and can challenge police legitimacy, as the use of predictive software may undermine the ability of individual officers or police authorities to account for their decisions (Bennett Moses and Chan, 2018). Digitalization could also create more self-fulfilling prophecies, as problems and harmful situations may arise as police bodies act against predicted futures (Bowling et al., 2019: 33, see also Fyfe et al., 2018). Thus, digitization changes intelligence, and has also extended the powers of the police.

Conclusion

Intelligence is an elusive and dynamic concept. Intelligence-led policing emerged in the context of supporting traditional investigative police-led strategies and specialist police operations but is now increasingly framed as a process model used to underpin the managerial organizational structures designed to improve future police work. Intelligence represents a certain type of *actionable* knowledge about a specific criminal environment, intended to inform and reduce uncertainties for decision-makers (Fyfe et al., 2018). For criminology it is essential to understand and explore what becomes the focus of intelligence when intelligence issues are related to crime, deviance, and risk. New trends in risk and crime control can influence intelligence practices, as can the development of digital technologies. Developments in intelligence are closely linked to technological advancements, as they provide new ways of collecting, storing, analyzing, and presenting data and information. These advancements also link intelligence analysis to other digital technologies, such as prediction technologies and other data integration platforms (see Platforms by Egbert).

Main takeaways

- Intelligence is an elusive concept but has a long history in both governmental and non-governmental organizations aiming at 'knowing the enemy,' often using secret methods.
- Intelligence can be understood both as a process, a product, a managerial concept, and as a practice.
- Intelligence is not neutral and objective but reflects the political context it works within.
- Digitalization has opened for new ways of collecting, storing, processing, and sharing intelligence.
- *Who* and *what* is defined as a threat and risk emerge from intelligence practices and is influenced by the digital tools and systems used throughout the process.

Suggested reading

Fyfe, N. R., Gundhus, H. I., & Rønn, K. V. (2018). *Moral Issues in Intelligence-led Policing.* London: Routledge.

Gill, P., & Phythian, M. (2018). *Intelligence in an Insecure World* (3rd ed.). Cambridge: Polity Press.

Ratcliffe, J. H. (2016). *Intelligence-led Policing* (2nd ed.). London: Routledge.

Sanders C. B., Weston, C., & Schott, N. (2015). Police innovations, 'secret squirrels' and accountability: Empirically studying intelligence-led policing in Canada. *British Journal of Criminology*, 55(4), 711 – 729.

Sheptycki J. (2017). The police intelligence division-of-labour. *Policing and Society*, 27(6), 620 – 635.

References

Adams, C., & Thompson, T. L. (2016). *Researching a Posthuman World: Interviews with Digital Objects.* London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-57162-5.

Bennett Moses, L., & Chan J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing & Society,* 28(7), 806 – 822. https://doi.org/10.1080/10439463.2016.1253695

Bowling, B., Reiner, R., & Sheptycki, J. (2019). *The Politics of the Police* (5th ed.). Oxford: Oxford University Press. https://doi.org/10.1093/he/9780198769255.001.0001

boyd, d. (2011). Social network sites as networked publics. Affordances, dynamics, and implications. In Z. Papacharissi (ed.), *A Networked Self: Identity, Community, and Culture on Social Network Sites* (pp. 39 – 58). New York: Routledge.

Brayne, S. (2020). *Predict and Surveil. Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Brodeur, J.-P. (1983). High and low policing: Remarks about the political activities. *Social Problems*, 30(5), 507 – 520.

Brodeur, J.-P. (2010). *The Policing Web.* Oxford: Oxford University Press.

Buckley, J. (2013). *Managing Intelligence: A Guide for Law Enforcement Professionals.* New York: CRC Press. https://doi.org/10.1201/b15515

Burcher, M., & Whelan, C. (2019). Intelligence-led policing in practice: Reflections from intelligence analysts. *Police Quarterly*, 22(2), 139 – 160. https://doi.org/10.1177/1098611118796890

Carter, J. G. (2013). *Intelligence-led Policing: A Policing Innovation.* El Paso: LFB Scholarly.

Chan, J., Sanders, C., Bennett Moses, L., & Blackmore, H. (2022). Datafication and the practice of intelligence production. *Big Data & Society*, 9(1). https://doi.org/10.1177/20539517221089310

Clemente, D. (2014). Cybersecurity. In R. Dover, M. Goodman, & C. Hillebrand (eds.), *Routledge Companion to Intelligence Studies* (pp. 256 – 263). London: Routledge.

Diderichsen, A. (2016). Om efterretningstjenesters legitimitet. In K. V. Rønn (ed.), *Efterretningsstudier* (pp. 67 – 88). Frederiksberg: Samfundslitteratur.

Ellefsen, B., & Lomell, H. M. (2024). Polisiær etterretning i et historisk perspektiv. In E. B. Unneberg, P. Jansen, & O. Trønnes (eds.), *Etterretningsanalyse i politiet* (pp. 36 – 56). Oslo: Universitetsforlaget.

Ericson, R. V., & Haggerty, K. D. (1997). *Policing the Risk Society.* Oxford: Clarendon Press. https://doi.org/10.3138/9781442678590

Flyverbom, M. (2022). Overlit: Digital architectures of visibility. *Organization Theory*, 3(3). https://doi.org/10.1177/26317877221090314

Fyfe, N. R., Gundhus, H. I., & Vrist Rønn, K. (2018). *Moral Issues in Intelligence-led Policing.* London: Routledge.

Garland, D. (2001). *The Culture of Control.* Oxford: Oxford University Press.

Gill, P., & Phythian, M. (2018). *Intelligence in an Insecure World* (3rd ed.). Cambridge: Polity Press. https://doi.org/10.1111/1478-9302.12073_52

Gundhus, H. I. (2013). Experience or Inowledge? Perspectives on new knowledge regimes and control of police professionalism. *Policing: A Journal of Policy and Practice*, 7, 176 – 192.

Gundhus, H. O. I., Talberg, N., & Wathne, C. T. (2022). From discretion to standardization: Digitalization of the police organization. *International Journal of Police Science & Management*, 24(1), 27 – 41. https://doi.org/10.1177/14613557211036554

Hillebrand, C. (2014). Intelligence oversight and accountability. In R. Dover, M. Goodman, & C. Hillebrand (eds.), *Routledge Companion to Intelligence Studies* (pp. 305 – 312). London: Routledge.

Innes, M., & Sheptycki J. (2004). From detection to disruption: Intelligence and the changing logic of police crime control in the United Kingdom. *International Criminal Justice Review*, 14(1), 1 – 24. https://doi.org/10.1177/105756770401400101

Innes, M., Fielding, N., & Cope, N. (2005). 'The appliance of science?': The theory and practice of crime intelligence analysis. *The British Journal of Criminology*, 45(1), 39 – 57. https://doi.org/10.1093/bjc/azh053

James, A. (2013). *Examining Intelligence-Led Policing. Developments in Research, Policy and Practice.* Basingstoke: Palgrave Macmillan.

Kaufmann, M. (2023). *Making Information Matter: Understanding Surveillance and Making a Difference.* Bristol: Bristol University Press.

Kaufmann, M., Egbert, S., & Leese, M. (2018). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692. https://doi.org/10.1093/bjc/azy060

Kent, S. (1966). *Strategic Intelligence for American World Policy.* Princeton: Princeton University Press.

Lundgaard, J. M., & H. O. I. Gundhus (2024). Den digitale hviskeleken: Data, informasjon og kunnskap i etterretningsstyrt politiarbeid. In E. B. Unneberg, P. Jansen & O. Trønnes (eds.), *Etterretningsanalyse i politiet* (pp. 220 – 232). Oslo: Universitetsforlaget.

McCulloch, J., & Wilson, D. (2016). *Pre-crime. Pre-emption, Precaution and the Future.* Abingdon: Routledge.

O'Neil, C. (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy.* New York: Crown.

Ratcliffe, J. H. (2016). *Intelligence-led Policing* (2nd ed.). London: Routledge. https://doi.org/10.4324/9781315717579

Reiner, R. (1985). *The Politics of the Police.* Brighton: Harvester Wheatsheaf.

Rønn, K. V. (2016). Efterretningsetik udenfor lovens rammer. In K. V. Rønn (ed.), *Efterretningsstudier* (pp. 219 – 246). Frederiksberg: Samfundslitteratur.

Sanders C. B., Weston C., & Schott N. (2015). Police innovations, 'secret squirrels' and accountability: Empirically studying intelligence-led policing in Canada. *British Journal of Criminology*, 55(4), 711 – 729.

Sheptycki, J. (2004). Organizational pathologies in police intelligence systems: Some contributions to the lexicon of intelligence-led policing. *European Journal of Criminology*, 1(3), 307 – 332. https://doi.org/10.1177/1477370804044005

Stout, M., & Warner, M. (2018). Intelligence is as intelligence does. *Intelligence and National Security*, *33*(4), 517 – 526. https://doi.org/10.1080/02684527.2018.1452593

Vestby, A. (2018). Policy-making without politics. Overstating objectivity in intelligence-led policing. In N. R. Fyfe, H. O. I. Gundhus, & K. V. Rønn (eds.), *Moral Issues in Intelligence-led Policing* (pp. 265 – 282). Routledge. https://doi.org/10.4324/9781315231259-14

Warner, M. (2013). The past and future of the Intelligence Cycle. In M. Phythian (ed.), *Understanding the Intelligence Cycle* (pp. 9 – 20). London: Routledge.

Weston, C., Bennett-Moses, L., & Sanders, C. (2019). The changing role of the law enforcement analyst: Clarifying core competencies for analysts and supervisors through empirical research. *Policing and Society*, 30(5), 532 – 547. https://doi.org/10.1080/10439463.2018.1564751

Sanja Milivojevic

34 Internet of things

Abstract: IoT systems are poised to play a critical role in today's infrastructures. These developments, though slow-paced, are followed by concerns about privacy violations and pervasive surveillance that bring new crime opportunities. IoT devices have not been designed with privacy and social implications in mind; they are developed using a technical, not sociotechnical, approach to technology development. This narrow approach is mainly responsible for vulnerabilities of systems, as well as often clumsy legal and policy attempts to 'catch up' with technology. As a result, we can expect further quests to 'patch up' glitches in technology's application and the unanticipated adverse impacts of technology on individuals and communities. Research must go beyond the usual frame of identity theft, blackmail, hacking, and profiling to explore the impact of the IoT on domestic and family abuse, intimate partner violence, and digital forensics. Community-led, interdisciplinary methodologies are helpful platforms for imagining and designing alternative futures in digital criminology, including the IoT.

Keywords: the IoT, privacy, surveillance, sociotechnical approach, interdisciplinary methodologies

The Internet of Things (IoT) has captured the attention of global media outlets, businesses, consultancy companies, researchers from STEM (science, technology, engineering, mathematics) disciplines, and to a lesser extent, social sciences in the last two decades. Whether warranted or not, some authors call the IoT systems "the next technological revolution" (Blythe et al., 2020), a ground-breaking but also disrupting innovation, particularly in the context of the so-called 'smart city' (see Smart City by Hayward). The smart city represents a novel way of managing urban environments by connecting and monitoring critical infrastructure components and services via emerging technologies (e. g., artificial intelligence—AI, machine learning, cloud computing, blockchain, data analytics, the IoT). In the smart city, the IoT plays a vital role in transport and traffic management, pollution monitoring, resource and energy management, healthcare, manufacturing, and the like, by collecting and exchanging data about the environment, optimizing services, and in doing so, enhancing the quality of life for smart city's inhabitants. At the same time, the IoT developments and deployments are followed by concerns about privacy violations and pervasive surveillance via networked objects (Maras and Wandt, 2019).

The IoT refers to the ability of everyday objects with a unique identification number and Internet Protocol addresses (which refers to the object's *unique identity*) to connect to the internet, other communication networks, and each other (*network*) via machine-to-machine communication (M2M), often without direct human intervention. Such objects generate and exchange data via sensors (*sensing and communication*, actuation capability that can capture context and provide/exchange information about

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-035

the thing itself or its environment) and act based on data analysis. Thus, IoT technologies can be recapped as objects with unique identities, connected in a network, with the capability to sense, communicate, and perform specific actions.

The IoT's promise lies in enhancing the quality of services provided to people and improving the quality of life by harnessing the power of sensors, connectivity, data processing, and user interfaces (Saini et al., 2022). Ultimately, these systems enable the automation of tasks and save time and money for people, companies, and government agencies (Maras and Wandt, 2019). One example of the IoT is the so-called 'smart security' integrated system that often includes objects such as glass and door break detectors, motion detectors, smart doorbells, security alarms, and the like. Smart heating of homes and offices is another well-known example of the IoT systems.

Definitions and applications

There are many typologies of the IoT. By application, they can be consumer, commercial, industrial, military, and infrastructure systems. They can be autonomous or conventional (the IoT that requires human involvement and analysis) and centralized or decentralized. As mentioned above, the technology is already deployed in homes and buildings (security systems, smart TV, speakers, thermostats, lighting, home assistants, doorbells), energy networks (intelligent grid with resource optimization, monitoring real-time consumption, improving distribution within the supply chain, and sustainability), vehicles and roads (smart navigation, traffic prediction, sensors, remote software updates, predictive maintenance), and humans (wearables, smart watches, health trackers, heart implants, glasses).

While IoT has been considered the 'next big thing' in technology for quite some time (Blythe and Johnson, 2021), the technology has yet to 'cross the chasm' and become ubiquitous. Many, including the author of this entry, have been blindsided by the 'guesstimates' about the growth of this type of M2M systems. Such guesstimates were predicting that, for example, in the year 2032, each person is likely to be connected to 3,000–5,000 everyday things within the IoT (Milivojevic and Radulski, 2020) and that there will be more than 125 billion devices worldwide by 2030 (Riley, 2023). Just how these prophecies are being made, based on what methodology and how they seep into the academic and common knowledge are fundamental questions that, unfortunately, I must leave behind because of the word count limitation for this entry. In saying this, I would like to encourage critical contributions on this issue as a priority.

Nevertheless, we see such forecasts in almost every industry, government, or academic publication, even though, as I mentioned above, projections such as that the average UK household will have 15 internet-connected products by 2020 have been slow to materialize. The slower-than-expected pace occurred mainly because of significant privacy and security concerns caused by inadequate security provisions, such as weak default passwords, bad encryption, and lack of software updates (Blythe et al.,

2020; Blythe and Johnson, 2021). All this is not new; market innovations, products, and services are rarely designed with security or, more broadly, with social implications in mind. They are often considered as a technical innovation alone, a mathematical and engineering problem that needs to be solved via mathematical or engineering methods and equations, with very little, if any, consideration of the social aspect of such advances in the design segment of the innovation process. This linear technological approach to technology design and production is, I argue, largely responsible for vulnerabilities of systems and legal and policy attempts to 'catch up' with its deployment. Such an approach also results in never-ending quests to 'patch up' glitches in technology's application and, more worryingly, the unanticipated adverse impact technology has on individuals and communities.

The IoT crime, intelligence and surveillance: the problem, the solution, and how the solution creates a (different) problem

Scholars have highlighted the IoT's relevance for criminology since the early 2010s, stressing the potential of the technology to generate new crime opportunities, such as burglary, stalking, and sex crimes (Blythe and Johnson, 2021). As Felson and Eckert note in *Crime and Everyday Life*, technology can "leap ahead and create problems, while nonmaterial culture lags behind the technology – unable at first to address those problems" (2023: xvi). Delays in attitudes, legislation, and regulation are particularly obvious in debates around artificial intelligence (AI), especially large language and generative AI models such as OpenAI's ChatGPT (see Artificial Intelligence by Van Brakel). The IoT managed to avoid this lag due to a slower pace of technological development in this field.

The IoT's key weaknesses are linked to security and privacy, making them vulnerable to cyberattacks (Saini et al., 2022). It is often argued that the devices lack security provisions by default (Blythe et al., 2020), as they are not developed with security in mind. Given that the adoption of the technology is growing, scholars in the field of crime science suggest that the attack surface is also expanding (Johnson et al., 2022). Think, for example, about your smart doorbell (e.g., Amazon Ring) or your smart TV. Did you change the default password the device comes with? Do you even know how to enhance your security protocols? I confess I do not. Even if we do, it is often impossible to upgrade the security, as little care and attention are given to devices' crime and security implications (Johnson et al., 2022). Many IoT systems do not have a privacy policy that you sign before use, or at least they are challenging to find (Maras and Wandt, 2019). Finally, manufacturers often provide very little publicly available information about the security features of their devices (Blythe et al., 2019). Having this in mind, potential cyberattacks on networked devices can be local, for example when hackers exploit security weaknesses in the device to gain un-

authorized access to the device or the network (see Hacking by Wall). They can also be broader/global, when machines are 'hijacked' and used as botnets to carry out medium or large-scale attacks. In 2016, in one of the first global IoT attacks, hackers exploited hundreds of thousands of internet-connected devices with cameras and video recording systems. They launched Distributed Denial of Service (DDoS) attacks, shutting down Twitter and Netflix, among thousands of other businesses (Blythe and Johnson, 2021).

Another potential avenue for offenders is committing data or privacy breaches, in which people's personal data, such as personal identifiers and health information, could be exposed via the IoT. For example, a compromised microphone or video camera can allow a hacker to listen in on audio within a room or obtain the visual of the surroundings, allowing access to private conversations and other data which can be particularly important in the context of government or industry espionage (Maras and Wandt, 2019). In this example, the IoT's vulnerability to cyberattacks has been intertwined with its capacity for harmful surveillance. Yet, one must dig deep to find actual cases of IoT-based offending, other than often sensationalist news reports and the above-mentioned global-scale attacks. In this field, lab-based research—research that tests vulnerability in a simulated environment—dominates the space, pointing out potential exposures and gaps, which, it is suggested, offenders might use in the future (Blythe and Johnson, 2021; Maras and Wandt, 2019). As such, empirical research on the scale of victimization and offending, locally and globally, is warranted (see Victimization by Walklate).

In the studies of IoT-related offending academia focused traditionally on identity theft (see ID theft by Langford, Wærstad, and Svensson), blackmail, hacking into the IoT devices, and profiling—gathering data from the IoT about household occupancy to orchestrate theft (Blythe and Johnson, 2021). The role of the IoT in domestic and family abuse, intimate-partner violence, and stalking has been sporadically investigated in social sciences, including criminology (Slupska and Tanczer, 2021). Most of the existing literature on the topic has been published by STEM researchers in STEM journals, advocating, often, technological solutions for the problem, such as defensive devices like panic buttons, emergency lockets, safety bags, self-defensive jackets, intelligent bracelets, bands, and other wearables, even smart sandals and IoT robots. Google Scholar's results for 'IoT' and 'violence against women' display a range of papers from STEM disciplines with above-proposed solutions. The absence of social sciences' contribution is puzzling, given the attention paid to incidents in which the IoT systems are used as a tool of violence. Examples are discussed in a report from one of the leading technology magazines, *Wired*, on how smart tech is used for domestic abuse (July 2018), a report from technology portal *Vice* on abusers exploiting smart home devices (October 2019) and a UK parliament report on technology and domestic abuse (November 2020).

Some examples reported in the media, albeit extreme, underline the criminological relevance. In 2018, *The New York Times* reported that survivors of IoT-facilitated abuse were often subjected to dystopian activity, such as air conditioners being remotely switched off, digital front door passcodes changed every day, boiling a kettle of water to let the victims know they are being watched, and doorbells ringing without anyone

being outside (Lo, 2021). In one of the first documented court cases involving IoT abuse from 2018, an individual was convicted of stalking his estranged wife by hacking into the smart home hub and using it as a gateway to listen to her conversations and stalk her. Following this case, the Gender and Internet of Things project at University College London has concluded that there is both a lack of data on technology abuse via this technology and that tech abuse via the IoT (or tech abuse in general) is not explicitly considered in risk assessments and safety plans (Lopez-Neira et al., 2019). Authors of one of the few studies looking at the technology-facilitated abuse via the IoT correctly pointed out that such abuse is a sociotechnical issue that requires reframing in research design and approach, including technology co-creation and co-design (Slupska and Tanczer, 2021). I come to this important point at the end of this contribution.

Yet, the 'crime harvest' (see Johnson et al., 2022)—that is a wave of offending that was poised to follow the adoption of the IoT, both in terms of threats to individual consumers and access to larger critical infrastructure—is yet, if ever, to happen. Nonetheless, these narratives are carefully crafted by cybersecurity experts from big tech, who, after pointing out the 'problem'—the vulnerability of the original product or a lack of updates or security 'patches' in connected devices—offer a solution. Such solutions often consist of buying protection against the threat, preferably their own product. An example is a post on CNBC from January 2023, in which cybersecurity experts from Google-owned cybersecurity firm Mandiant pointed out IoT as the 'next big hacking prize,' only to offer the security solution by Mandiant. Other Big Tech cybersecurity experts followed suit (cf. 'The dark web's criminal minds see the Internet of Things as next big hacking prize', *CNBC Cyber Report*, 9 January 2023).

With the IoT's capacity for data aggregation, exchange, and analysis, privacy and mass surveillance issues are important caveats. As IoT devices record and transmit a staggering amount of data about users' activities, preferences, habits, purchases, and routines, some authors argue that the level of surveillance users and bystanders are subjected to is so unprecedented that it has previously been written about only in science fiction novels (Maras and Wandt, 2019). What presents a challenge to societies concerned with privacy and data protection is that IoT devices are not run by any government, but by multiple private entities.

This caveat, of course, does not mean that government agencies are not hoping to access and benefit from big data collected by the IoT systems. Law enforcement agencies worldwide have been seeking and obtaining data from smartwatches, medical devices, smart assistants (Alexa, Siri, etc.) and other IoT-based systems, successfully utilizing them in criminal investigations (Maras and Wandt, 2019). A particular use of the IoT has been identified in digital forensics, or the 'Internet of Evidence' (Milivojevic and Radulski, 2020). Recently, researchers at the University of Lausanne in Switzerland conducted experiments in which they found that IoT gadgets might be especially helpful in investigating arson. Experiments have shown that the devices, including temperature and motion sensors, smart cameras, smoke detectors, and a voice assistant, aided in establishing the precise moment/time of the fire, the room where the fire started, and enabled a reconstruction of the timeline of events (Choi, 2023). Some authors

(cf. Hany Atlam, Ahmet Alenezi, Maria Stoyanova, Aine MacDermott, Thar Baker, Shancang Li) now suggest that IoT Forensics is a new branch of digital forensics with a wide range of potential sources of evidence, such smart devices, servers, personal area networks, and cloud data.

Conclusion

IoT systems follow the familiar pattern of other technological advancements: they are considered within the 'part of the problem' and 'part of the solution' narrative. The IoT devices have not been designed with privacy and social implications in mind; they are developed using the technical, not socio-technical, approach to technology development. As this contribution demonstrates, there are many guesstimates and unknowns when it comes to this technology that requires criminologists' immediate attention.

Main takeaways

- The IoT brings new crime opportunities. Although its development has been slower than, for example, advances in AI, the IoT's key weaknesses (security and privacy) are making them vulnerable to cyberattacks and data and privacy breaches.
- The role of IoT in identity theft, blackmail, hacking and similar crimes is solidly researched (particularly by the STEM and crime science branch of criminology). This is not the case with the IoT's role in family and domestic violence.
- While offending via the IoT is not widespread, there are many 'solutions' advocated by a range of commercial actors ('solving the problem with the money' approach). Academia should be cautious about encouraging such 'solutions.'
- Surveillance via the IoT is ubiquitous and performed by a range of private actors. Government agencies are increasingly tapping into the big data generated by the IoT, hoping to gather intelligence necessary to solve criminal cases.

Future engagements

- Criminology should engage with the IoT systems, both as a potential target for offending and a tool for crime control. Some applications of the IoT in offending need more scrutiny, such as family and domestic violence. Privacy concerns and surveillance performed by the IoT need to be examined, both in terms of potential offending and crime control.
- Criminologists should not research the IoT alone, and not only in the field of crime science, as has been the case in the last twenty years. In order to grasp dynamics and challenges interdisciplinary approaches as different expertise is needed, including criminology, law, computer science, and engineering to venture past a

fragmented depiction of the much more complex picture of technological innovation. Interdisciplinarity is necessary to understand the complexities, risks, and potential remedies for the problems identified in this space.
– Empirical interdisciplinary research in digital futures is both sporadic and lacking. Lab-based research and literature reviews are helpful, but criminology needs to set new foci on understanding the nature, characteristics, and relevance of the IoT to crime, offending, victimisation, and criminal justice interventions.
– Potential future developments should not be guessed. There is no single, predetermined, or certain future. Scholarly activity is needed to move beyond mere observation by participating in future(s) co-creation and co-design through the broader process of sociotechnical innovation, with value-in-design research from a range of fields. Community-led, interdisciplinary methodologies via innovative research collaboration spaces are a platform for imagining and designing alternative futures in digital criminology.

Through this interdisciplinary and empirical research glitches, lags, and unwanted consequences of new technologies can be addressed, including the IoT. By using a sociotechnical approach in researching crime and digital technologies, new ways of creating academic knowledge and technological innovations can be forged while having a meaningful, non-digital impact in this space.

Suggested reading

Blythe, J. M., & Johnson, S. D. (2021). A systematic review of crime facilitated by the consumer Internet of Things. *Security Journal*, 34(1), 97 – 125.

Johnson S. D., Blythe, J. M., Kim, E., & Sombatruang, N. (2022). Crime and the consumer Internet of Things. In M. Gill (ed.), *The Handbook of Security* (pp. 705 – 727). Cham: Springer International Publishing.

Maras, M.-H., & Wandt, A. S. (2019). Enabling mass surveillance: Data aggregation in the age of big data and the Internet of Things. *Journal of Cyber Policy*, 4(2), 160 – 177.

Milivojevic, S., & Radulski, E. M. (2020). The "future Internet" and crime: Towards a criminology of the Internet of Things. *Current Issues in Criminal Justice*, 32(2), 193 – 207.

Slupska, J., & Tanczer, L. M. (2021). Threat modelling intimate partner violence: Tech abuse as a cybersecurity challenge in the Internet of Things. In J. Bailey, A. Flynn, & N. Henry (eds.), *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*, Emerald Studies in Digital Crime, Technology and Social Harms (pp. 663 – 688). Leeds: Emerald Publishing Limited.

References

Blythe, J. M., & Johnson, S. D. (2021). A systematic review of crime facilitated by the consumer Internet of Things. *Security Journal*, 34(1), 97 – 125. doi:10.1057/s41284 – 019 – 00211 – 8

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. doi:10.1186/s40163 – 019 – 0110 – 3

Blythe, J. M., Sombatruang, N., & Johnson, S. D. (2019). What security features and crime prevention advice is communicated in consumer IoT device manuals and support pages? *Journal of Cybersecurity*, 5(1), tyz005. doi:10.1093/cybsec/tyz005

Choi, C. (9 May 2023). *The Internet of Things: Fire Sleuth, Fire Starter.* IEEE Spectrum. spectrum.ieee.org/iot-for-arson-forensics, accessed 19 June 2023.

Felson, M., & Eckert, M. (2023). *Crime and Everyday Life: A Brief Introduction*, 6th edition. Thousand Oaks: Sage. doi:10.4135/9781506394817

Johnson, S. D., Blythe, J. M., Kim, E., & Sombatruang, N. (2022). Crime and the consumer Internet of Things. In M. Gill (ed.), *The Handbook of Security* (pp. 705 – 727). Cham: Springer International Publishing. doi:10.1007/978-3-030-91735-7_33

Lo, M. (2021). A domestic violence dystopia: Abuse via the Internet of Things and remedies under current law note. *California Law Review*, 109(1), 277 – 316.

Lopez-Neira, I., Patel, T., Parkin, S., Danezis, G., & Tanczer, L. (2019). "Internet of Things": How abuse is getting smarter. *Safe – The Domestic Abuse Quarterly*, 63, 22 – 26. doi:10.2139/ssrn.3350615.

Maras, M.-H., & Wandt A. S. (2019). Enabling mass surveillance: Data aggregation in the age of big data and the Internet of Things. *Journal of Cyber Policy*, 4(2), 160 – 177. doi:10.1080/23738871.2019.1590437

Milivojevic, S., & Radulski, E. M. (2020). The "future Internet" and crime: Towards a criminology of the Internet of Things. *Current Issues in Criminal Justice*, 32(2), 193 – 207. doi:10.1080/10345329.2020.1733452

Riley, A. (2023). *How your smart home devices can be turned against you.* BBC Future. bbc.com/future/article/2020511-how-smart-home-devices-are-being-used-for-domestic-abuse, accessed 19 June 2023.

Saini, S., Chauhan, A., Sapra, L., & Thakur, G. (2022). Emerging trends in security, cybercrime, and digital forensics in the context of the Internet of Things. In K. Kaushik, S. Dahiya, A. Bhardwaj, & Y. Maleh (eds.), *Internet of Things and Cyber Physical Systems* (pp. 1 – 22). Boca Raton: CRC Press.

Slupska, J., & Tanczer, L. M. (2021). Threat modeling intimate partner violence: Tech abuse as a cybersecurity challenge in the Internet of Things. In J. Bailey, A. Flynn, & N. Henry (eds.), *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*, Emerald Studies in Digital Crime, Technology and Social Harms (pp. 663 – 688). Leeds: Emerald Publishing Limited. doi:10.1108/978-1-83982-848-520211049

Catherine Adams and Terrie Lynn Thompson

35 Interviews with digital objects

Abstract: 'Interviewing objects' is a concept and method of qualitative research inquiry that includes nonhuman entities as active participants. It highlights the co-constitutive roles of humans and technologies in shaping social and cultural formations. Employing heuristics from posthuman theoretical perspectives, it documents and analyzes interactions between humans and nonhuman objects like digital technologies. The approach provides insights into the dynamic interdependencies and contributions of nonhuman entities to everyday practices and societal changes.

Keywords: posthuman inquiry, human–nonhuman interaction, Actor-Network Theory, postphenomenology, relational ontology

Introduction

'Interviewing objects' was coined as a counterpoint to the popular anthropocentric method of interviewing (human) subjects in qualitative inquiry. The phrase signals the intent to include not only human subjects but also *nonhuman* objects as active participants in human and social science research. Nonhuman objects are any material entities involved in a crime or law enforcement event, for example, a gun, credit card, or Apple AirTag, as well as software such as a digital app, facial recognition system, or social media site. In practice, 'interviewing objects' describes deploying a wide range of practical methods or 'heuristics' borrowed, adapted, and sometimes invented to aid posthuman and more-than-human research inquiry. The heuristics assist the researcher in attuning to the world of things, documenting human–nonhuman interconnections and relational correspondences, and analyzing the contributions of nonhumans to everyday practices, societal change, and cultural formations.

The word interview is from the Middle French *s'entrevoir* which means "to see one another, to meet" but also "to see indistinctly or in passing" ('Interview', 2023). While 'interviewing' is commonly understood as having a face-to-face meeting with a person, interviewing an object recalls a now obsolete definition which is "to look at (a person or thing); to glimpse or glance at (a person or thing)." Thus, in conducting an object interview, the researcher endeavors to catch glimpses of the object in action as it performs and mediates the gestures and understandings of its (human) employer, as well as involved others and things. In fact, an object interview is about interviewing human–nonhuman (subject–object) relational associations or assemblages as they unfold in everyday practice. For example, Actor-Network Theorist, Latour (1999) once asked: "Who or what is responsible for the act of killing … the citizen or the gun?" (p. 178). His answer? "*Someone else*"! (p. 179). Latour's *someone else* is a new hybrid actor: a citizen-gun, gun-citizen, or simply, a gunman. This hybrid relational configuration alters both entities; a person becomes a different subject by wielding the gun, just

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-036

as the gun assumes a new role when held by a human, signaling a dynamic interdependency where each entity is transformed by the other's influence. To interview an object thus entails finding opportunities to observe, document, and analyze its active participation in everyday interactions with humans and other nonhuman entities.

Interviewing objects has special relevance for digital criminology. Researchers must increasingly grapple with the material contributions of digital technologies in the commission of and the investigation of contemporary crimes. Wood et al. (2023) point out that digital—and non-digital—technologies have the capacity to not only extend criminal agency and amplify harm but also *mediate* human experience and action in manifold ways (see Agency by Krasmann). Indeed, "it is impossible to think about a crime that does not involve any kind of materiality. Because who would the hacker be without a keyboard? Or the robber without a weapon?" (Hermansen, 2020: 24). To account for digital technologies' co-constitutive, co-agential, and mediating roles in crime studies, some researchers are turning to approaches such as Actor-Network Theory (ANT) and postphenomenology for assistance (van der Wagen, 2019; Wood et al., 2023). Such inquiries engage a variety of posthuman, more-than-human, and new materialist methods and theories to interview these newly admitted nonhuman research participants.

Interviewing object heuristics

As an approach to doing digital research, interviewing objects describes the deployment of one or more heuristics derived from posthuman theoretical positions and philosophical traditions, including Actor-Network Theory (ANT), postphenomenology, media ecology, and more recently, new and feminist materialisms, agential realism, and Indigenous theories of nonhuman agency. What binds these diverse philosophies and theories together is a shared commitment to a relational ontology beyond human-centric, Euro-Western substantive ontologies. Relational ontologies recognize that humans and nonhuman things generate and share agency in and through their relationships in the ongoing, lived contexts of their social, cultural, and historical worlds.

In *Researching a Posthuman World: Interviews with Digital Objects*, Adams and Thompson (2016) present a set of eight heuristics for interviewing and 'speaking with' nonhuman things. Earlier versions of the heuristics appeared in Adams and Thompson (2011) and Thompson and Adams (2013). The heuristics deploy "specific tricks" to coax digital things "to offer descriptions of themselves, to produce scripts of what they are making others—humans or non-humans—do" (Latour, 2005: 79). The first four focus on 1) attuning the researcher to digital things in the midst of their collaborations and correspondences with humans (and other nonhumans) and 2) generating posthuman qualitative materials needed for analysis:

1. *Gathering anecdotes*

   The researcher constructs 'anecdotes' or short, descriptive accounts of specific events as they unfolded, and illuminate how digital objects interact with humans

in everyday life. These narratives are detailed, specific, and reflective of real-life experiences, and focus on describing what transpired rather than explaining why. A posthuman anecdote aspires to 'reassemble and resemble' human interactions with digital objects, and as such, can provide insight into the co-constitutive, co-agential dynamics between humans, nonhumans, and their environmental situation. The primary 'interview' question here is: "Describe how the [digital] object or thing appeared, showed up, or was given in professional practice or everyday life. What happened?" (Adams and Thompson, 2016: 24)

2. *Following the actors*
   This heuristic encourages the researcher to get a sense of which actors matter as a specific practice is being enacted. In other words, who or what is acting, what are they doing, and how do they go about it? This is often a fun aspect of object interviews as one traces the moves and movements of both human and non-human actors. Keep in mind that the aim of this work is not to just generate a list of actors but rather to begin to get a glimpse of how these actors are assembling: Who-what seems to be the more powerful actors? Who-what is sidelined? Other heuristics will enable you to further untangle the negotiations and relations between actors and how they come together, mingle, animate, fall apart, and dissolve in an array of assemblages.
3. *Listening for the invitational quality of things*
   This heuristic recognizes that nonhuman things 'speak' to humans and to each other in unique and powerful ways. How a nonhuman's 'invitation' (suggestion, nudge, and even command) is 'heard' or perceived by a human depends on the affordances or possibilities for action or interaction offered by the digital object (Affordances by Wood and Arpke-Wales), as well as on the intentions or orientation of the human to their world. Here the researcher attends to the specific invitations a digital technology issues to its user, and the silent conversations and gestural actions that subsequently unfold between humans and nonhumans in practice.
4. *Studying breakdowns, accidents, and anomalies*
   As a technology is incorporated into our everyday practices, it tends to slip seamlessly into the background of our lives. But when that technology malfunctions, behaves unpredictably, or is unexpectedly missing, the taken-for-granted object suddenly leaps into focus. Posthuman researchers are especially alert to such 'breakdown' moments since they not only make the digital object in question apparent but, importantly, illuminate the dynamic web of human–nonhuman relations supporting the practice (before it was disrupted).

Having attuned to digital presences and absences and gathered preliminary posthuman data, the next four heuristics provide possible ways 1) to gently loosen and analyze the human–nonhuman entanglements at the research site and 2) to reveal otherwise hidden aspects of a digital object's participation in events:

1. *Discerning the spectrum of human–technology relations*
   In his postphenomenological analyses of different technologies, Don Ihde (1990)

uncovers different types of human–technology–world (HTW) relations structuring our everyday actions: embodiment, hermeneutic, alterity, and background. Embodiment relations extend and enhance our body's abilities (e.g., an automobile extends our legs). Hermeneutic relations position technology as an interpretive intermediary, reshaping and refining our understanding of reality (e.g., a thermometer translates hotness and coldness into a number on a scale). Alterity relations describe our interaction with technology as a distinct 'other' (e.g., a chatbot), while background relations acknowledge technology's silent shaping of our environment, influencing us without direct engagement (e.g., central heating). Understanding these basic types of posthuman/more-than-human interactions is valuable for critically examining how different technologies influence our actions, perceptions, cognition, and sociality in manifold ways.

2. *Applying the laws of media*
   Media ecologists understand technologies as environments that mediate and participate in the creation and transformation of society and culture. McLuhan and McLuhan (1988) devised a popular analytic tool, the 'laws of media,' to probe the breadth of a given technology's radiating effects and side effects on societies. The laws provide four more interview object questions: What does a technology enhance? What does it obsolesce? What does it retrieve from the past? What does it reverse into when we abuse or over-rely on it?
3. *Unravelling translations*
   Originating in ANT, this heuristic examines how different actors come to be powerful through specific actions with other actors. Focusing on translations enables the researcher to study the ways in which human and nonhuman actors come together (including how they enroll each other), how they negotiate and manage their interactions, what sorts of networks or assemblages take shape (or fall apart), and how these actor-networks change. Here the researcher dives into the politics of assemblages: in other words, paying attention to what actors do, their impacts on other actors, and the kinds of actions and ideas which circulate. Attention focuses on the multiple realities and assemblages that come into view. This includes 'collateral realities' (Law, 2011), that is, the unintended realities and the gaps between practices and what realities are enacted. The tensions in coexisting, not-always-harmonious networks can offer rich insights into sociotechnical practices and researchers will usually find multiple moments of translation (see Translation by Wilson-Kovacs).
4. *Tracing responses and passages*
   Breathing new life into Latourian actor networks, Tim Ingold (2012) reconceives the field of action as a "meshwork," that is, as a "co-responsive movement of occurrent things along their manifold lines of becoming" (p. 437). Here the researcher traces "the flows of energy and circulations of materials" (Ingold, 2012: 417) as actors thread their way through their world, "co-responding" with the things around them, wayfinding and improvising new passages. The process creates new meshworks. Attention shifts to the living ecology of "things thinging" (Heideg-

ger, 1971), which includes the movements and gestures of people and things corresponding with one another. Foregrounded is how entities and practices come to be and how things and people "become" in these practices.

Adams and Thompson's (2016) eight heuristics offer possible starting points for including digital things as meaningful research participants. Each provides a different sensitivity for recovering nonhuman contributions at the qualitative research site. Not all heuristics need to be applied; diverse situations and practices demand different heuristics or the invention of new approaches. Since these eight heuristics are drawn from different theoretical origins, overlaps, as well as contradictions, are expected. Thus, researchers need to become more familiar with the concepts and theoretical lineages behind these heuristics as they engage with them (see Adams and Thompson, 2016, and other references above). Much like a semi-structured qualitative interview, the posthuman researcher must be ready to adapt their methods in the field to "glimpse" the digital object of interest in its "dance of agency" (Pickering, 2013: 78) with humans.

Consider the complexity of criminology practices. There is a vast number of digital actors that are implicated in the perpetration of crime, assist in investigating criminal activities, are employed to enable citizen and community-based crime prevention, and are used to address justice and social justice agendas. Body-worn videocams, hate-based text and images, surveillance and tracking devices, predictive policing algorithms, Facebook, recidivism statistics, witness statements, ransomware, bots, and data dashboards are entangled with citizens, police, case workers, criminologists, victims, offenders, and IT specialists in places such as the street and in homes, courtrooms, police stations and vehicles, forensics labs, borders, and board rooms. Digital technologies encompass devices, networks, apps, code, algorithms, analytics, and increasingly sophisticated forms of AI and machine learning. The material saturation of digital criminology practices demands theoretical and conceptual frameworks and sensibilities that can make visible and critically reckon with how human and digital actors co-constitute the everyday practices, frame thinking, and ways of being in this particular field. Adams and Thompson (2016) position digital things as co-participants in research projects and even co-workers; a notion that resonates with van der Wagen's (2019) assertion that researchers should regard things as more important and active participants in studies of cybercrime.

A brief example. Wood et al. (2023) do not specifically reference 'interviewing objects,' but their work illustrates how posthuman heuristics may be mobilized in criminological research. Variations of '*Listening for the invitational quality of things*' and '*Unravelling translations*' (two heuristics listed above) are used by Wood et al. to construct a new framework for understanding how technology mediates violence. They show how attending to a technology's invitations assists the researcher to "analyze the particular way that the technology has contributed to conducting violence instead of functioning merely as a vehicle for that violence" (2023: 8). They also consider "the role design decisions might play in co-producing harm" (p. 8) to aid the identification of "strategies for ameliorating technology-facilitated violence that address the specific

technological element producing such harms" (p. 10). Such an approach could be added to the toolbox of heuristics for interviewing digital objects.

Wood et al. (2023) adapt Latour's notion of translation to clarify that "harm *translation* produces meaning rather than merely *transporting* it" (p. 11, italics in original). Wood et al.'s adaptation underlines the mediating, co-constitutive relationship between actors and the digital objects they employ. Recalling the basic amplification–reduction structure of all HTW relations (cf. Ihde, 1990), digital objects are understood to "alter the experiential and connotative landscape invoked in contemplating and pursuing these ends, amplifying certain experiential qualities whilst simultaneously reducing others" (Wood et al., 2023: 11).

Implicit in Wood et al.'s framework is the need for criminologists also to gather detailed posthuman anecdotes (Heuristic 1) when studying technology-mediated harm. In Wood et al. (2023), posthuman anecdotes appear only in roughed-out form (e.g., "Apple AirTags being used for unwanted tracking or Find My iPhone being used to track someone's movement surreptitiously" (p. 9)). Of course, the article is a theoretical proposal for reconceptualizing technology-mediated violence rather than a case study of a particular digitally enhanced criminal occurrence. The latter would require documented "detail [of] the causative power technology can exert in *co-producing* and *inducing* harms (p. 2, italics in original) and "accounting for the various relations that might exist between intentions (both users' and designers'), ends and technologies involved in the co-production of any harmful event" (p. 3).

Indeed, gathering and generating posthuman anecdotes is central to the research practice of interviewing objects. Without these more-than-human "reassembled resemblings" (Adams and Thompson, 2016: 31) of humans and nonhuman things in the entangled flow of their correspondences and collaborations, the researcher can only speculate on how the digital may be contributing to, mediating, and co-constituting technology-enhanced crimes and investigative practices. Posthuman anecdoting requires writerly practice combined with an orientation to the technology-saturated world that begins with noticing, wondering, and posing questions of and with the digital things involved in everyday events. In this way, "micro and macro analysis is meshed together to make visible layers, multiplicities, movements, the politics of assemblages, and presences/absences" of the togetherness of human-digital working and living (Thompson, 2023: 233). Such inquiry is a form of "slow research" (Law and Singleton, 2012): a deliberate lingering with the puzzles that emerge as the researcher untangles the flow of human and nonhuman assemblings.

Moving forward

Digital criminology research imperatives are evolving, driven by increased awareness of the powerful collaborations forged between human and nonhuman entities. The focus is not merely on the rising adoption of digital technologies but on understanding the profound changes these technologies are facilitating, mediating, and co-enacting.

The scholarly discourse is shifting from a deterministic, anthropocentric point of view to one that embraces more intricate, embodied, and material relationships and co-constitutions. More-than-human theories and methods offer new ways to include digital devices, data, and systems—from tracking devices and text messages with hate speech to body cameras and predictive algorithms—and explore their co-active role in shaping criminal activities as well as control and policing practices.

In this context, there is a unique potential for establishing new sociotechnical configurations, diverse political dialogues, and multiple realities. Focusing on the co-relational interplay of humans, objects, and systems could lead to a fundamental rethinking of what we consider 'real' and how we engage with it. Object interview methods offer a useful tool to scrutinize and disrupt digital narratives, proposing alternative ways of imagining our sociotechnical world. Similarly, the sprawling data infrastructures that govern our visibility and memory can also be reimagined (see Infrastructures by Grisot and Parmiggiani), countering the overly optimistic techno-utopian narratives that often dominate the conversation.

Object interviewing stands as a part of the larger postqualitative inquiry landscape aiming to bolster the credibility and political implications of more-than-human-oriented studies. Thompson and Adams (2020) have recently suggested three guiding principles to assess the quality of such inquiries. These include: 1) explaining or making evident how the researcher speaks *with* things, including critically questioning how their own digital practices affect the researcher and the work of research; 2) weaving and fusing human and nonhuman storylines; and 3) acknowledging the liveliness of posthuman research work to bring about different questions, practices, and therefore, realities. These dynamics are not meant as a definitive checklist of what constitutes a good posthuman/more-than-human account but rather to open a space for continuing to discuss these questions. They are also very good points to make explicit when writing about one's research, including methodology sections.

Conclusion

Interviewing objects is an approach to doing digital criminological inquiry that explicitly includes nonhuman entities as active research participants. It provides a set of eight heuristics or starting points for criminology researchers to 'interview' posthuman/more-than-human assemblages, e. g., the 'citizen-gun, gun-citizen' and other heterogeneous gatherings of humans and nonhumans. Such an approach "forces [the researcher] to abandon the subject-object dichotomy" (Latour, 1994: 34) and rethink the instrumental perspective that technologies are neutral objects.

Interviewing objects:

– Recognizes the co-constitutive relationships between humans and technology and the digital's amplifying effects in criminal activities as well as control and policing practices.

– Unsettles and shifts the gaze of the researcher beyond just human participants, enabling them to attend to, describe, and analyze complex human–digital assemblages.
– Affords the researcher opportunities to examine technology-supported practices as well as the morality, ethics, and politics of digital–human relations.

Suggested reading

Adams, C., & Thompson, T. L. (2016). *Researching a Posthuman World: Interviews with Digital Objects.* London: Palgrave Macmillan.

Thompson, T. L., & Adams, C. (2020). Accountabilities of posthuman research. *Explorations in Media Ecology*, 19(3), 337 – 349.

Van der Wagen, W. (2019). The significance of "things" in cybercrime: How to apply Actor-Network Theory in (cyber)criminological research and why it matters. *Journal of Extreme Anthropology*, 3(1), 152 – 168.

Verbeek, P. P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things.* Chicago: University of Chicago Press.

References

Adams, C., & Thompson, T. L. (2011). Interviewing objects: Including educational technologies as qualitative research participants. *International Journal of Qualitative Studies in Education*, 24(6), 733 – 750. https://doi.org/10.1080/09518398.2010.529849

Adams, C., & Thompson, T. L. (2016). *Researching a Posthuman World: Interviews with Digital Objects.* London: Palgrave Macmillan.

Heidegger, M. (1971). *Poetry, Language, Thought* (A Hofstadter, trans.). New York: Harper & Row

Hermansen, E. (2020). *How a border told became a border controlled: From narratives to material solutions – developments in EU migration control.* [Master's thesis] University of Oslo, Oslo. https://www.duo.uio.no/handle/10852/81463

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth.* Bloomington: Indiana University Press.

Ingold, T. (2012). Toward an ecology of materials. *Annual Review of Anthropology*, 41, 427 – 442. Doi: 10.1146/annurev-anthro-081309 – 145920

'Interview'. (2023, July). In *Oxford English Dictionary*, Oxford University Press. https://doi.org/10.1093/OED/3888688934

Latour, B. (1994). On technical mediation. *Common Knowledge*, 3(2), 29 – 64.

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies.* Cambridge, MA: Harvard University Press.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network Theory.* Oxford: Oxford University Press.

Law, J. (2011). Collateral realities. In R. F. Dominguez & P. Baert (eds.), *The Politics of Knowledge* (pp. 156 – 188). London: Routledge.

Law, J., & Singleton, V. (2012). ANT and politics. Working in and on the world. http://www.sv.uio.no/sai/english/research/projects/newcomers/publications/working-papers-web/ant-and-politics.pdf

McLuhan, M., & McLuhan, E. (1988). *Laws of Media: The New Science.* Toronto: University of Toronto Press

Pickering, A. (2013). Being in an environment: A performative perspective. *Natures Sciences Sociétés*, 21(1), 77 – 83. https://doi.org/10.1051/nss/2013067

Thompson, T. L. (2023). Researching with, on, in and through the postdigital: Accounting for more-than-humanness. In P. Jandrić, A. MacKenzie, & J. Knox (eds.), *Postdigital Research. Postdigital Science and Education* (pp. 231 – 247). Cham: Springer.

Thompson, T. L., & Adams, C. (2013). Speaking with things: Encoded researchers, social data, and other post-human concoctions. *Distinktion: Scandinavian Journal of Social Theory*, 14(3), 342 – 361. https://doi.org/10.1080/1600910X.2013.838182

Thompson, T. L., & Adams, C. (2020). Accountabilities of posthuman research. *Explorations in Media Ecology*, 19(3), 337 – 349. https://doi.org/10.1386/eme_00050_7

Van der Wagen, W. (2019). The significance of "things" in cybercrime: How to apply Actor-network Theory in (cyber)criminological research and why it matters. *Journal of Extreme Anthropology*, 3(1), 152 – 168. http://dx.doi.org/10.5617/jea.6895

Wood, M. A., Mitchell, M., Pervan, F., Anderson, B., O'Neill, T., Wood, J., & Arpke-Wales, W. (2023). Inviting, affording and translating harm: Understanding the role of technological mediation in technology-facilitated violence. *The British Journal of Criminology*, 63(6), 1384 – 1404. https://doi.org/10.1093/bjc/azac095

Stefano Mazzilli Daechsel

36 Labs

Abstract: The laboratory is an important but largely understudied phenomenon in criminology. This chapter delves into four criminologically relevant aspects of laboratories: (i) labs and the production of criminological knowledge, (ii) forensic labs and criminal investigations, (iii) labs as sites of crime and harm, and (iv) labs as sites of resistance. If labs are particularly pertinent to digital criminology, it is because they are increasingly digitalized and networked spaces, affecting their functionalities and relations to wider society.

Keywords: laboratory, knowledge production, forensic science, crime, resistance

Introduction

Like many academic disciplines, criminology was born in the laboratories of 19th-century researchers who pioneered a set of practices and technological constellations that formed the spatial, temporal, and material backbone of the modern sciences. Cesare Lombroso and the positivist school of criminology's physiological theories of deviance may have fallen out of favor due to their reliance on deeply flawed methods, but the laboratory persists as a site of criminological knowledge production in contemporary biosocial approaches. From forensic labs to clandestine drug labs, laboratories are also crucial to both criminal investigations and criminal organizations. This chapter delves into four criminologically relevant aspects of laboratories: (i) labs and the production of criminological knowledge, (ii) forensic labs and criminal investigations, (iii) labs as sites of crime and harm, and (iv) labs as sites of resistance. If labs are particularly pertinent to digital criminology, it is because they are increasingly digitalized and networked spaces (see Digital by Wernimont). Special attention will be paid to this process of digitalization and how it has affected the laboratory's many functionalities and its relationship to wider society.

Disclaimers and acknowledgments: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 947681).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-037

Labs and the production of criminological knowledge

While it is "almost impossible to imagine science without laboratories" (Schmidgen, 2021: 1), it is all too easy to forget that criminology emerged from the laboratories of 19th-century criminologists who searched for the causes of deviance in the measurement and visualization of 'criminal' bodies (Lombroso, 2006). This legacy of lab work continues to inform a great deal of criminological research, particularly in the subfields of biosocial criminology (Rafter et al., 2016) and neurocriminology (Nordstrom et al., 2011). Laboratory experiments are frequently touted as examples of scientifically rigorous, evidence-based criminology because they ostensibly offer researchers greater control over the conditions under which certain variables are studied (Nagin and Pogarsky, 2003; Yang et al., 2009). These studies are often published in high-ranking journals such as *Criminology* and they give the discipline the veneer of a hard science.

Biosocial criminology and neurocriminology have been criticized for overemphasizing individual and biological factors in explaining criminal and deviant behavior, to the detriment of environmental and societal factors that, by definition, cannot be studied in a lab setting (Fallin et al., 2019; Burt, 2023). Despite these critiques, little attention has been paid to the lab as a site of knowledge production in criminology. There is, however, a rich and diverse body of ethnographic literature on laboratories in the field of science and technology studies (STS), including Bruno Latour and Steve Woolgar's (1979) seminal work *Laboratory Life.*

Based on extensive fieldwork in a neuroendocrinology laboratory, Latour and Woolgar (1979: 105) argue that a laboratory should be regarded "as a system of literary inscription, an outcome of which is the occasional conviction of others that something is a fact." Scientific facts are *constructed* through a slow and practical process 'by which inscriptions are superimposed and accounts backed up or dismissed' (Latour and Woolgar, 1979: 236). This reduction of laboratory practices to a series of operations on literary inscriptions proved controversial and garnered its fair share of critics (e. g., Tilley, 1981; Murphy, 1994; Hacking, 1999). Regardless, *Laboratory Life* inaugurated the subfield of laboratory studies and inspired many STS scholars to conduct laboratory ethnographies (e. g., Knorr Cetina, 1981; Lynch, 1985; Bowker and Star, 1999; Doing, 2009). Criminologists working under the umbrella of digital criminology would do well to engage with this STS literature to better understand the role of labs in the production of criminological knowledge.

Forensic labs and criminal investigations

The role of forensic or crime laboratories in the functioning of law enforcement has been explored from a variety of angles. In criminology, much of this work has focused on the unrealistic impressions that laypeople might have of forensic scientists and their

technological capabilities. Highly fictionalized depictions of forensic practices and technologies in popular media, epitomized by the American television series *Crime Scene Investigations (CSI)*, were feared to have a real and detrimental effect on the functioning of the criminal justice system by distorting juror expectations of the kind of forensic evidence required for a criminal conviction. The '*CSI* effect', as it came to be known, provoked a lot of media and scholarly debate over its validity and even about the effect that media coverage of the phenomenon might have on criminal justice outcomes (Cole and Dioso-Villa, 2009). Meanwhile, academics and public institutions such as the American National Research Council have sounded the alarm over the lack of quality control and sufficient institutional independence from law enforcement in many forensic labs, arguing that "with the exception of nuclear DNA analysis, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source" (National Research Council, 2009: 7).

Despite these concerns about crime labs and the recognition of their highly consequential role in the outcomes of the criminal justice system, there are but a few ethnographic studies of crime labs that examine the day-to-day practices of forensic scientists and their changing technological infrastructures (see Infrastructures by Grisot and Parmiggiani).[1] Two noteworthy studies are Corinna Kruse's (2016) *The Social Life of Forensic Evidence* and Beth Bechky's (2020), *Blood, Powder and Residue.* Kruse traces the production of forensic evidence as it moves from site to site within the Swedish criminal justice system. She dedicates a chapter to Sweden's only forensic laboratory (the Swedish National Laboratory of Forensic Science), emphasizing the probabilistic logics that forensic scientists employ to manage the uncertainty inherent to their work (Kruse, 2016: 70).

Bechky (2020), meanwhile, relates the findings of an 18-month ethnographic study of a crime lab in the United States. She details the practices by which "messy crime scenes are transformed into clean scientific reports and courtroom evidence" (Bechky, 2020: 7–8) and, crucially, how these practices are undergoing transformations due to technological and organizational innovations. These innovations include the shift away from forensic generalists who could perform most tasks in the lab (toxicology, firearms, fingerprint analysis, etc.) to highly specialized practitioners, resulting in increasingly fragmented labs in which units have "distinct social practices" (Bechky, 2020: 19). She also addresses the forensic scientist's relationship to digital technologies, especially algorithmic tools intended to automate, supplement, or even replace human expertise, arguing that "these experts work in dialogue with their instruments and techniques, and their embodied and community understandings are needed for interpretation and translation" (Bechky, 2020: 182). Here again we have an opportunity for digital criminology to build on this work and narrow in on the incorporation of digital

1 For books aimed at a more general audience on forensic labs and the many problems in forensic science, see Garrett (2021) and Murphy (2015).

technologies in forensic practices, such as DNA phenotyping (Kaufmann and Vestad, 2023), to understand how these new technologies affect the politics of knowledge production in forensic labs and criminal investigations.

Labs as sites of crime and harm

Laboratories are also sites of harmful and/or criminal behavior. Popular media is rife with representations of clandestine drug labs and, thanks in part to the success of television series like *Breaking Bad*, they are typically associated with the production of methamphetamine in the US. In addition to being illegal, drug labs can pose enormous health risks to the people working in them and are often sites of labor exploitation (Chiu et al., 2011). These laboratories speak to the historical and etymological links between laboratories and workshops. The Latin term *laboratorium* originally referred to a task or work, but around the 15th century it came to denote the "workshops of alchemists, apothecaries and metallurgists" (Schmidgen, 2021: 4) who required controlled and secluded spaces to practice their crafts. Clandestine drug labs retain elements of the pre-modern laboratory, with its mixture of scientific practices, material production, and secrecy.

The relationship between harm and laboratories extends well beyond clandestine drug labs. Harm, abuse, and exploitation have been all-too-common features of scientific experimentation in lab settings. Unethical experimentation on humans might seem a thing of the past in an era of ethics committees and strict regulation, but this ignores a more complicated reality. To evade ethical oversight, "experimentation sponsored by the pharmaceutical industry has progressively migrated to the [Global] South" (Negri, 2017: 1022), resulting in some grotesque abuses of clinical trial participants. One particularly egregious case was the American pharmaceutical company Pfizer's testing of an experimental antibiotic on Nigerian children in 1996, which resulted in the death of eleven children and major bodily injury to many others (Ezeome and Simon, 2008). If we widen our lens to consider the abuse of animals in laboratory settings, then the amount of harm perpetrated in labs in the name of science or consumer product development increases exponentially (Goyes and Sollund, 2018).

The digital dimension of lab-based harm and criminality warrants greater scrutiny. On the one hand, the expanding use of computer simulation and big data systems (see Big Data by Završnik) might have the effect of reducing the need for unethical experimentation and facilitate the sharing of experiment data (to the benefit of lab rats across the world). On the other hand, the digitization of consumer products has given profit-driven actors greater capacities to exploit consumers through personal data collection. What's more, the software embedded in smart products can be designed to circumvent legal restrictions, as was the case with Volkswagen's 'defeat devices' in their vehicles, i. e., software that altered the car's nitrogen oxide emissions in a test setting to give the impression that it complied with emission restrictions when in fact emissions far surpassed legal limits under normal driving conditions (Fitzgerald and Spencer,

2020). What is clear is that digital criminology has an opportunity to make meaningful contributions to the study of deviance in technological and scientific communities and examine how harmful practices are embedded in the pursuit of scientific knowledge (Ben-Yehuda, 1986).

Labs as sites of resistance

Finally, labs can also act as sites of resistance to hegemonic power structures. Hackerspaces, makerspaces, and fabrication laboratories (Fab Labs) have emerged in recent decades as networked spaces where people interact with digital technologies in new and sometimes subversive ways. My own work has focused on the maker movement, an international network of people who make things using digital fabrication machines such as 3D printers and laser cutters, and how the movement might be considered a form of resistance to consumerism (Mazzilli-Daechsel, 2019). Makers, as those who belong to the movement call themselves, tend to conduct their activities in makerspaces or Fab Labs where members have access to digital fabrication equipment and, just as importantly, the opportunity to collaborate with or learn from other members. Projects are sometimes documented and then shared on databases and wikis online. Fab Labs will also organize events for the general public to promote technological literacy. This combination of in-person and online community building is a good example of how digital and networked technologies can transform the nature and functions of labs, from enclosed and exclusive spaces of knowledge production to networked and more inclusive spaces of experimentation, creativity, and alternative forms of material production.

The case study of the maker movement should also temper any overenthusiastic appraisal of lab-based activism or resistance. The fact remains that makerspaces are biased in their design towards highly skilled users and can be difficult to navigate for the uninitiated (see Bias by Oswald and Paul). The majority of makers already have an educational or professional background in technology or design, and tend to be male and middle-class, meaning that the opportunities that makerspaces and Fab Labs afford their users are flowing to a relatively homogenous demographic, one that already has a relatively high degree of technological literacy. Much like public libraries assume a certain level of literacy across the population it serves, makerspaces are dependent on a certain baseline of technological competence held by a mostly privileged minority. Digital criminology should continue to study the intersections of political resistance and laboratories where they arise, such as in citizen science initiatives and climate change activism.

Conclusion

The laboratory is an important but largely understudied phenomenon in criminology. As I have tried to demonstrate, it can serve as a site of many criminologically relevant practices. The growth in lab studies in STS offers a host of methodological and conceptual tools that should be taken up by criminologists to better understand the role of labs in crime, crime control, and the discipline of criminology itself. For digital criminology, it is important to explore the ways in which digital and networked technologies are changing the functions and capabilities of specific labs or networks of labs, paying particular attention to the ways in which processes of digitalization challenge "the iconography of the laboratory as an isolated, self-contained structure in which equally isolated scientists sit in front of a microscope" (Schmidgen, 2021: 24).

From my own experiences studying labs, and by way of conclusion, here are some dynamics to consider when conducting an ethnographic study of a laboratory:

- The interactions between humans and humans, humans and nonhumans (machines and animals), as well as between nonhumans and nonhumans, how they shape the spatial arrangement of the lab and the temporal rhythm of its operations;
- The relationships between inside and outside. Labs have an internal coherence, but they are always connected to several outsides, be they wider societal structures, illicit economies, personal histories, or digital networks;
- The production of subjectivities within the lab. As Schmidgen (2021: 24) reminds us, laboratories don't just produce new knowledge and technologies, they also produce "new types of people."

Suggested reading

Bechky, B. (2020). *Blood, Powder, and Residue.* Princeton, NJ: Princeton University Press.

Kruse, C. (2016). *The Social Life of Forensic Evidence.* Oakland, CA: University of California Press.

Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts.* London: Sage Publications.

Schmidgen, H. (2021). The laboratory. *Encyclopedia of the History of Science.* doi: 10.34758/sz06-t975

References

Bechky, B. (2020). *Blood, Powder, and Residue.* Princeton, NJ: Princeton University Press.

Ben-Yehuda, N. (1986). Deviance in science: Towards the criminology of science. *British Journal of Criminology*, 26(1), 1 – 27.

Bowker, G., & Star, S. (1999). *Sorting Things Out: Classification and its Consequences.* Cambridge, MA: MIT Press.

Burt, C. (2023). Irreducibly social: Why biocriminology's ontoepistemology is incompatible with the social reality of crime. *Theoretical Criminology*, 27(1), 85 – 104.

Chiu, Y., Leclerc, B., & Townsley, M. (2011). Crime script analysis of drug manufacturing in clandestine laboratories: Implications for prevention. *British Journal of Criminology*, 51, 355 – 374.

Cole, S., & Dioso-Villa, R. (2009). Investigating the 'CSI Effect' effect: Media and litigation crisis in criminal law. *Stanford Law Review*, 61(6), 1335 – 1373.

Doing, P. (2009). *Velvet Revolution at the Synchrotron: Biology, Physics and Change in Science.* Cambridge, MA: MIT Press.

Ezeome, E., & Simon, C. (2008). Ethical problems in conducting research in acute epidemics: The Pfizer meningitis study in Nigeria as an illustration. *Developing World Bioethics*, 10, 1 – 10.

Fallin, M., Whooley, O., & Barker, K. (2019). Criminalizing the brain: Neurocriminology and the production of strategic ignorance. *BioSocieties*, 14, 438 – 462.

Fitzgerald, A., & Spencer, D. (2020). Governmentality and environmental rights: Regulatory failure and the Volkswagen emissions fraud case. *Critical Criminology*, 28, 43 – 63.

Garrett, B. (2021). *Autopsy of a Crime Lab: Exposing the Flaws in Forensics.* Oakland, CA: University of California Press.

Goyes, D., & Sollund, R. (2018). Animal abuse, biotechnology and species justice. *Theoretical Criminology*, 22(3), 363 – 383.

Hacking, I. (1999). *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Kaufmann, M., & Vestad, M. (2023). Biology and criminology: Data practices and the creation of anatomic and genomic body 'types'. *Critical Criminology*, 31, 1217 – 1232.

Knorr Cetina, K. (1981). *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science.* Oxford: Pergamon Press.

Kruse, C. (2016). *The Social Life of Forensic Evidence.* Oakland, CA; University of California Press.

Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts.* London: Sage Publications.

Lombroso, C. (2006) *Criminal Man* (M. Gibson & N. Rafter, Trans.). Durham, NC: Duke University Press.

Lynch, M. (1985). *Art and Artifact: A Study of Shop Work and Shop Talk in a Research Laboratory*, London: Routledge.

Mazzilli-Daechsel, S. (2019). Simondon and the maker movement. *Culture, Theory and Critique*, 60(3 – 4), 237 – 249

Murphy, E. (2015) *Inside the Cell: The Dark Side of Forensic DNA.* New York: Nation Books.

Murphy, R. (1994). The sociological construction of science without nature. *Sociology*, 28, 957 – 974.

Nagin, D., & Pogarsky, G. (2003). An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity. *Criminology*, 41(1), 167 – 194.

National Research Council. (2009). *Strengthening Forensic Science in the United States: A Path Forward.*

Negri, S. (2017). Unethical human experimentation in developing countries and international criminal law: Old wine in new bottles? *International Criminal Law Review*, 17, 1022 – 1048.

Nordstrom, B., Gao, Y., Glenn, A., Peskin, M., Rudo-Hutt, A., Schug, R., Yang, Y., & Raine, A. (2011). Neurocriminology. *Advances in Genetics*, 75, 255 – 283.

Rafter, N., Posick, C., & Rocque, M. (2016). *The Criminal Brain, Second Edition: Understanding Biological Theories of Crime.* New York: NYU Press.

Schmidgen, H. (2021). The laboratory. *Encyclopedia of the History of Science.* doi: 10.34758/sz06-t975

Tilley, N. (1981). The logic of laboratory life. *Sociology*, 15, 117 – 126.

Yang, Y., Raine, A., Narr, K., Colletti, P., & Toga, A. (2009). Localization of deformations within the amygdala in individuals with psychopathy. *Archives of General Psychiatry*, 66, 986 – 994.

Maja Vestad

37 Low-tech

Abstract: We should be careful of overstating the importance of technologies that inspire dramatic narratives. All digital tools and practices have material counterparts, and while some may be less flashy, shiny, and indeed less captivating, they are key players in the fields that we study. This chapter discusses how high- and low-tech operates in tandem and suggests that low-tech must be taken seriously in the development of digital criminological theory.

Keywords: materiality, technology, crime scene investigation, surveillance, observation

Introduction

The technologies through which we tell our stories influence how we view the world. Within forensics, for instance, many different methods and tools (bite mark analysis, lie detectors, blood pattern analysis, to name some) have been hailed as 'the ultimate tool,' 'the most accurate' instrument—as saviors, as solutions to the many problems of establishing truths about crimes and about people (see e. g., Bowers, 2019; Bunn, 2012). What is often forgotten, however, is the complex landscape of other technologies that such methods and tools emerge from. While new digital technologies tend to capture our attention, the epistemic and social cultures in which they exist also consist of more mundane technologies and practices. As researchers, we can acquire knowledge that is valuable for new and emerging digital practices by moving our focus away from the most shiny, flashy, and captivating technologies. For example, in her work *The Carrier Bag Theory of Fiction* (2019), Ursula Le Guin posits that the story of human evolution has predominantly revolved around violence, propelled by the fascination with weapons as technologies that endowed early humans with means to defend, attack, and eat. However, she writes, the carrier bag was probably "the first cultural device … a container to hold gathered products and some kind of sling or net carrier" (Fisher, 1975 in Le Guin, 2019). Carrier bags enabled humans to collect food to bring back to the group, to carry children and belongings across space. An exploration of this development, as opposed to the narrative of the weapon, would tell a more collaborative

Disclaimers and acknowledgments: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 947681).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-038

and indeed feminine story of the becoming of technological society. I draw on this analogy to say that we should be careful of overstating the importance of technologies that inspire dramatic narratives. What we choose as our vantage point in research influences what kind of knowledge we produce.

Technologies are tools. As per the drive of the technological evolution, new tools replace old tools and that which was new becomes old. As such they are continuations of each other and exist across a spectrum of technological development. The term *low-tech* refers to technological tools and methods that are different from high-tech. If high-tech is advanced, expensive, and complex, low-tech is an umbrella term for technologies that are more manual, sustainable, cost-effective, and require less expertise to understand and operate. I have through fieldwork in forensic evidence communities examined more closely the ways in which technological tools frequent in the production of evidence and knowledge about crime.

Forensic geneticists, for instance, have intricate and careful procedures for handling and analyzing DNA traces, as do crime scene investigators. When a trace is located, the key aim is to figure out to whom it belongs and how it got there, and different technological tools are used by both expert groups in the process. Highly sensitive DNA detection machines may be used to scan an item in search for a trace. Software programs are then used to visualize information about DNA profiles so that they are easier to read by machines and non-machines (humans), to compare results, and make some estimations about the appearance and genealogy of the DNA owner (see DNA/Big Genome Data by Kaufmann). An ID number is generated to avoid confusion with other samples. Forensic evidence production is in a high-tech perspective a digitized process occurring within the capacities of software systems, digital storage, algorithms, and the internet.

What, then, if we shift our focus to low-tech in the same process? Paper bags with cotton swabs or confiscated items such as clothing and weapons arrive at the lab via regular post alongside a physical paper, which details the request of the police and includes some information about the investigation. The items are examined individually on a desk behind doors that regulate air pressure, by people in protective suits and two sets of plastic gloves. The items and the traces then need storage (a room) in the same way that the digitalized trace needs storage (a database). The room needs an oversight system and items must be kept apart to avoid contamination. From my observations, these high and low technologies appear to operate in tandem, in a sort of forensic ecosystem of material and less-material entities and capacities. An acknowledgment of this confluence allows for a greater understanding of the relationship between the analog and the digital.

Including a discussion on low-tech practices in a book on digital criminology is important as it can be argued that any digital information in contemporary societies has its foundation in some form of low-tech and material entity. Recognizing and understanding these roots is essential for a comprehensive understanding of the interconnected relationship between analog and digital realms and how they influence each other. Moreover, by using low-tech as a starting point, we can acquire a more profound

understanding of contemporary advancements within digital technologies. If we agree that technologies are key to the production of knowledge, then we may come to a better understanding of the world by studying how they work in their contexts. This chapter makes a case for not losing focus on the role of low-tech in relation to high-tech advancements. First, low-tech is conceptualized relative to high-tech, before the role of low-tech in early, modern, and contemporary criminological theory is described. Finally, the ideological properties of low-tech are discussed, both within and outside of criminology.

The relationship between high-tech and low-tech

The duality of high and low suggests that they are in some way mutually dependent and as opposites on a spectrum of technological complexity. If low-tech is the 'less smart' version of something newer, it can only exist once higher-tech has been developed. Take the example of communication technology. When phones were first introduced to the police, they were large permanent fixtures on street corners. Patrolling officers could walk over and make a call on the telephone if they wanted to report anything back to the station. Since then, of course, significant changes to phones have occurred—the police do not rely on street corner telephones to communicate any longer but have various digital methods for sharing information. A simple phone is not high-tech today. Instead, that which was once high-tech has since become low-tech through various levels of technological development and sophistication.

There is, of course, a danger in digital criminology to be obsessed with the latest digital device. By overstating the importance of new technologies, we risk understating the value of other tools and methods that accompany actors in the fields that we study, even if they may have more of an actual impact. High-tech depends on low-tech not only in the sense that low-tech had to be developed first for higher-tech to emerge, but every time we engage with high-tech, we are dependent on offline and analog technologies and environments that actively feed into higher technologies. As exemplified by my own fieldwork and outlined in the introduction, the accompanying letter that is sent along with the evidence through the mail holds crucial significance in enabling the capture and integration of DNA results into a data system. Not only are high and low methods and tools different from each other, but high-tech cannot function without low tech. Oftentimes data is collected and managed by offline technologies that are as simple as a paper form, and without these, data cannot be digitized, not fed into systems and not be analyzed by algorithms. That means high-tech and data always have an underlying concrete materiality and material infrastructure (see Kaufmann, 2017, 2023). Low-tech not only accompanies the fields that we study, but there is a relationship here that signifies an interdependent nature between the analog and the digital (see Digital by Wernimont), which we may miss by solely focusing on the newest device.

Studying how low-tech frequents high-tech environments may even give valuable insight into the attractions of high-tech. It helps challenging assumptions and biases

that may be embedded in high-tech solutions and cultures, allowing for a more inclusive understanding of technology and its societal impact. The point being, the way in which we orient our research influences our continued thinking about the field, including the development of theory.

High-tech and low-tech in the development of criminological theory

Many of the core theories in criminology were developed through the use of methods and tools that constituted cutting-edge technological advancements during their time. They were, however, accompanied by technologies that were less shiny but now used for new purposes—things like rulers, photographic cameras, and thermometers. This was at a time when criminologists worked with or on behalf of the police or criminal justice system to find 'root' causes of crime, which laid the foundation for the idea that crime can be studied scientifically (Rafter, 2004). As one of the early theorists, Alfonse Bertillon was the first person to systematize French arrest records to calculate recidivism rates in the 1880s. His innovative filing cabinet system made use of a combination of high and low-tech methods. It was pioneering in that he utilized photography (a high-tech at the time) for capturing the visual characteristics of arrested persons, and other anthropometric measuring tools such as rulers and weight scales to document other parts of their bodies. This system, compiled of both high and low technological methods, enabled the system to be ground-breaking and it was adopted by other Western countries (see Finn, 2017). The Italian school similarly embarked upon extensive documentation of arrested persons, employing a wide range of tools across the technological spectrum. One of Cesare Lombroso's main sources of information for his (now highly controversial) theories were the drawings, writings, and artwork of incarcerated people in Italy, which were analyzed alongside detailed measurements of their bodies (see Kaufmann and Vestad, 2023; Lombroso, 2006; see Labs by Mazzilli Daechsel). Both of these early schools of criminology lay some of the groundwork for future research into biosocial theories (see Rafter, 1997; Rose, 2000). Bertillon's system of identification later became a blueprint for racialized sentencing algorithms in the US (Wiggins, 2020). Core criminological theories developed in Chicago in the 1950s–1980s were similarly dominated by methods for generating data for analysis that were low-tech in their time—many were based on interview data. Through the use of analog technologies, they created a system for explaining and nuancing crime and control functions, i. e., drawing out some advantages and disadvantages of how the police, state, and prison system worked, how neighborhood dynamics influenced crime, and the effects of policy.

Major changes to technological capacities inspire new theoretical perspectives—a change that we see perhaps most clearly in surveillance studies. Significant technological changes have revolutionized the surveillance capacities of the state throughout his-

tory (see Coleman and McCahill, 2011). Historically, surveillance has encompassed "close observation, especially of a suspected person" (Marx, 2002: 8). According to Gary Marx, this definition "[made sense] in an age of servants listening behind closed doors, binoculars and telegraphic interceptions" (Marx, 2002: 10), but does not anymore, as human interactions and practical observation is increasingly replaced by data extraction and algorithmic predictions. Yet it is through these early forms of tech practices that the even-higher-tech mass surveillance could emerge (see Surveillance by Lyon). Further, high-tech data systems are needed to handle the masses of data collected, which means human interference and physical bodies 'disappear' (Lyon, 2001) from the act and field of surveillance, both in the sense that the observed becomes data points, and in the sense that those data points are analyzed by computers.

Yet even this high-tech environment is full of low-tech practices. For some local organized criminal networks, for instance, low-tech communication methods can be a way of remaining undetected by digital surveillance tools (Leukfeldt et al., 2017). Low-tech observation and intervention methods also remain at the core of intelligence collection as it is practiced by law enforcements globally, for instance at the French border, where migrants are often registered through paper-based methods that the police may utilize "to avoid that the digital trace would make them responsible of processing their asylum claim" while enabling refugees to tear them apart and "[erase] their legal and bureaucratic history in a certain country" (Tazzioli, 2023: 929). Bonelli and Ragazzi (2014) point to the importance of low-tech practices in counter-terrorism efforts by the French police. Meeting with informants and writing memos, from which arrest or deportation orders are later decided, are favored over high-tech network analyses as the most useful tool for "understanding the structure, the orientations, and the power relations within a group or organization" (2014: 489; see also Haggerty, 2012). Dahl and Svanaes (2020) show how the bodies of covert surveillance officers in Norway become equipped with a skill set of learned, tacit knowledge through years of practice, which enables them to maneuver their bodies for seeing and hearing while remaining largely unseen, similarly to the functioning of modern high-tech solutions such as CCTV cameras. Bozzini (2011) finds that the Eritrean state and military employ low-tech identification paperwork checks at road checkpoints as a means to instill fear in individuals contemplating evading mandatory military service. As these ID documents are not verified or cross-referenced with other data, this low-tech practice primarily serves the purpose of creating a perception amongst those being checked that they are under constant surveillance by a more powerful state, rather than actually logging their movements. These studies tell us that human interaction with low-tech continues to be part of the surveillance toolbox, both as means of observation and for remaining hidden. Studying the use of these technologies reveals something about their role in mediating relationships between actors, and about the means and aims of surveillance and security.

The ideological properties of low-tech

The appeal of low-tech is sometimes linked to a broader ideological stance against technologized society. Its use can represent a step away from the increasingly digitalized and automated workings of the world (see Automation by Mann), for which low-tech can operate as a counterbalance. Outside of criminology, we find initiatives advocating for a (re-)turn to a less technologically dependent society, to 'no-tech,' in which proponents "refuse to assume that each problem has a high-tech solution" (*No Tech Magazine*, 2023). Such initiatives promote a more natural and sustainable approach to technological innovation, which reduces resource consumption and is centered on human interaction and involvement (Tanguy et al., 2023). We find again this sentiment in studies that examine reluctances for adopting new methods and tools, also within law-enforcement. Some studies show that there is a multitude of reasons for why some law enforcement actors may resist implementing new advanced technological tools. Concerns over the consequences of organizational change, budget, or the potential loss of in-depth understandings of knowledge production processes are examples of what may draw actors in the police towards low-tech methods (Chan, 2001; Nhan and Huey, 2012; Vestad, 2024).

In my own work on crime scene investigation practices in Norway, I suggest that this narrative is not so straight-forward (Vestad, 2024). Rather, I find that low-tech is employed as a resource *additional to* high-tech practices, which enables investigators to triangulate results across methodologically different tools—essentially as a means to increase both accuracy and their own technological capacities. While new technologies such as sensitive machines for locating DNA traces have revolutionized the forensic field, low-tech tools, such as manual DNA analysis kits or sniffer dogs, remain key to the crime scene investigations process. Analog tools such as measuring tapes and drawing boards are used to document the layout of crime scenes and mark where trace materials have been collected from alongside virtual software. Using combinations of digital and analog tools enable investigators to gather information about traces while taking an active part in the knowledge production process. Low-tech remains as much a part of the process as high-tech for reasons other than techno-pessimism and challenges the notion that high-tech alone is the driving force of effective investigatory work. As such, the ideological properties of low-tech may even be instrumental in shaping technological development in forensic science.

Conclusion

It is important that criminologists continue to investigate the influences of technological change in the fields that we study. Low-tech may first appear irrelevant—something of the past, something from which the higher-tech we use today has emerged. Recent studies, however, show that low-tech is in continuous development and remains

central to practices across various fields, even (and perhaps especially) in digital high-tech environments. This tells us that criminological research on technology can benefit from also examining the roles of the less flashy and captivating technologies.

In conclusion, recognizing the significance of low-tech practices and their symbiotic relationship with high-tech is essential for a comprehensive understanding of digital criminology. All digital technologies are rooted in a material entity. By studying the interplay between analog and digital realms, we gain insights into the foundations of contemporary digital information that goes beyond the latest digital devices. Finally, by exploring the role of low-tech within high-tech environments, researchers can deepen their understanding of the field and contribute to the development of theory in digital criminology.

Main takeaways

- The interplay between high- and low-tech has been critical to the development of criminological theory.
- High-tech relies on low-tech, as all digital technologies are rooted in a material entity. Low-tech is crucial for enabling the capture, integration, and management of data, without which high-tech systems and analysis would not be possible.
- Digital criminology should be careful not to overstate the significance of the latest technological devices, as this can overshadow the importance of other tools and methods.

Suggested reading

Bonelli, L., & Ragazzi, F. (2014). Low-tech security: Files, notes, and memos as technologies of anticipation. *Security Dialogue*, 45(5), 476 – 493. https://doi.org/10.1177/0967010614545200.

Nhan, J., & Huey, L. (2012) 'We don't have these laser beams and stuff like that': Police investigations as low-tech work in a high-tech world. In S. Leman-Langlois (ed.), *Technocrime: Policing and Surveillance* (pp. 79 – 90). London: Routledge.

Tazzioli, M. (2023). Counter-mapping the techno-hype in migration research, *Mobilities*, 18(6), 920 – 935. DOI: https://doi.org/10.1080/17450101.2023.2165447.

References

Bonelli, L., & Ragazzi, F. (2014). Low-tech security: Files, notes, and memos as technologies of anticipation. *Security Dialogue*, 45(5), 476 – 493. https://doi.org/10.1177/0967010614545200.

Bozzini, D. M. (2011). Low-tech surveillance and the despotic state in Eritrea. *Surveillance & Society,* 9(1/2), 93 – 113. https://doi.org/10.24908/ss.v9i1/2.4102.

Bowers, C. M. (2019). Review of a forensic pseudoscience: Identification of criminals from bitemark patterns. *Journal of Forensic and Legal Medicine*, 61, 34 – 39. https://doi.org/10.1016/j.jflm.2018.11.001.

Bunn, G. C. (2012). *The Truth Machine: A Social History of the Lie Detector.* Baltimore: JHU Press.

Chan, J. B. L. (2001). The technological game: How information technology is transforming police practice. *Criminal Justice*, 1(2), 139 – 159. https://doi.org/10.1177/1466802501001002001.

Coleman, R., & McCahill, M. (2011). *Surveillance & Crime.* London: SAGE Publications Ltd.

Dahl, J. Y., & Svanaes, D. (2020). Hiding in plain sight: Directed surveillance as a bodily practice. *Surveillance & Society*, 18(4), 493 – 506. https://doi.org/10.24908/ss.v18i4.13555.

Finn, J. (2017). Making the criminal visible. In M. Brown & E. Carrabine (eds.), *Routledge International Handbook of Visual Criminology* (pp. 121 – 134). London: Routledge. https://doi.org/10.4324/9781315713281.

Fisher, E. (1975). *Woman's Creation: Sexual Evolution and the Shaping of Society.* New York: McGraw-Hill Book Co.

Haggerty, K. D. (2012). Surveillance, crime and the police. In K. Ball, K. D. Haggerty, & D. Lyon (eds.), *Routledge Handbook of Surveillance Studies* (pp. 235 – 244). London: Routledge.

Kaufmann, M. (2017). The co-construction of crime predictions: Dynamics between digital data, software and human beings. In N. R. Fyfe, H. O. I. Gundhus, & K. V. Rønn (eds.), *Moral Issues in Intelligence-led Policing* (pp. 143 – 160). London, New York: Routledge.

Kaufmann, M. (2023). *Making Information Matter. Understanding Surveillance and Making a Difference.* Bristol: Bristol University Press.

Kaufmann, M., & Vestad, M. (2023). Biology and criminology: Data practices and the creation of anatomic and genomic body 'types'. *Critical Criminology*, 31, 1217 – 1232. https://doi.org/10.1007/s10612-023-09732-6.

Le Guin, U. (2019). *The Carrier Bag Theory of Fiction.* London: Ignota Books.

Leukfeldt, E. R., Kleemans, E. R., & Stol, W. P. (2017). A typology of cybercriminal networks: From low-tech all-rounders to high-tech specialists. *Crime Law and Social Change*, 67, 21 – 37. https://doi-org.ezproxy.uio.no/10.1007/s10611-016-9662-2.

Lombroso, C. (2006 [1889]). Critical communication, in *Criminal Man* (M. Gibson& N. Rafter, trans.), (pp. 239 – 244). Durham, NC: Duke University Press.

Lyon, D. (2001). *Surveillance Society: Monitoring Everyday Life.* 1st ed. Buckingham and Philadelphia: Open University Press.

Marx, G. T. (2002). What's new about the "new surveillance"? Classifying for change and continuity. *Surveillance & Society*, 1(1). https://doi.org/10.24908/ss.v1i1.3391.

Nhan, J., & Huey, L. (2012). 'We don't have these laser beams and stuff like that': Police investigations as low-tech work in a high-tech world. In S. Leman-Langlois (ed.), *Technocrime: Policing and Surveillance* (pp. 79 – 90). London: Routledge.

No Tech Magazine. (2023). *About* [website]. Available: https://www.notechmagazine.com/about. Last accessed 24 July 2023.

Rafter, N. H. (1997). *Creating Born Criminals.* Champaign, IL: University of Illinois Press.

Rafter, N. (2004). The unrepentant horse-slasher: Moral insanity and the origins of criminological thought. *Criminology*, 42(4), 979 – 1008. https://doi.org/10.1111/j.1745-9125.2004.tb00542.x.

Rose, N. (2000). The biology of culpability: pathological identity and crime control in a biological culture. *Theoretical Criminology*, 4(1), 5 – 34. https://doi.org/10.1177/1362480600004001001.

Tanguy, A., Carrière, L., & Laforest, V. (2023). Low-tech approaches for sustainability: Key principles from the literature and practice. *Sustainability: Science, Practice and Policy*, 19:1. DOI: https://doi.org/10.1080/15487733.2023.2170143.

Tazzioli, M. (2023). Counter-mapping the techno-hype in migration research, *Mobilities*, 18(6), 920 – 935. DOI: https://doi.org/10.1080/17450101.2023.2165447.

Vestad, M. (2024). The persistent attractions of low-tech: Challenging the efficiency paradigm. *International Journal of Police Science and Management*, 26(2), 292 – 301. DOI: https://doi.org/10.1177/14613557241231164.

Wiggins, B. (2020). Crime. In *Calculating Race: Racial Discrimination in Risk Assessment, Religion and Democracy* (pp. 33 – 50). New York: Oxford Academic. https://doi.org/10.1093/oso/9780197504000.003.0003.

Avital Mentovich and Orna Rabinovich Einy

38 Online courts

Abstract: This chapter explores the transformation of courts from physical spaces to digital services. The rise of online courts, particularly through Online Dispute Resolution (ODR) and remote video proceedings, has redefined the legal landscape, leveraging technology to enhance access to justice. Initially focused on civil cases, ODR has expanded to include various types of disputes, facilitated by digital tools designed to streamline and democratize legal processes. However, the transition to online courts has also raised significant concerns about procedural fairness, particularly in criminal proceedings. Challenges related to the right to be present, the right to counsel, and the right to confront witnesses are exacerbated in virtual settings, with studies showing mixed results regarding the impact of remote proceedings on legal outcomes and perceptions of justice. The chapter critically examines these developments, highlighting both the opportunities and the potential pitfalls of integrating technology into the judicial process. As the use of online courts becomes more entrenched, the chapter calls for a nuanced understanding of how different mediums—written, audio, and video—affect various aspects of legal proceedings, from access to justice to the legitimacy of the legal system.

Keywords: online courts, access to justice, dispute resolution, legitimacy

Introduction

Are courts "a place or a service" (Susskind and Susskind, 2017)? To many of us, until quite recently, the answer was clear. Courts were synonymous with central, imposing buildings where synchronous proceedings took place, in the presence of the parties, before a judge. This began to change in the second decade of the 21st century, as several courts launched online proceedings, primarily in the civil context. These proceedings, often referred to as "online dispute resolution" or ODR (Katsh and Rifkin, 2001), involved the use of digital technology to facilitate remote communication, as part of formal and informal dispute resolution efforts (Katsh and Rabinovich Einy, 2017). Initially, these processes, which emerged in the e-commerce setting in the mid-1990s, were conducted through written asynchronous proceedings.

Use of ODR increased dramatically with the onset of COVID-19, in volume and variety of proceedings employed. Practically overnight, courts worldwide were forced to find solutions for the need to continue to hear and decide cases despite court closure

Disclaimers and funding: This research was supported by the Israeli Science Foundation, Grant No. 2381/22, "Online Courts in the Post-Covid Era: An Empirical Analysis." The views and opinions expressed in this chapter are, however, those of the authors alone and do not reflect those of the Foundation.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-039

and social distancing requirements (Engstrom, 2021; Sourdin et al., 2020). While use of written asynchronous ODR processes expanded, the vast majority of online proceedings used during the pandemic relied on videoconferencing (as well as audio) communication, which allowed for remote real-time interaction, with some or all parties communicating from afar. Millions of hearings, including various criminal proceedings, were conducted through what were now termed 'remote proceedings'—court processes conducted in writing, or through audio or video (Susskind, 2019). These developments raised significant challenges to traditional conceptions of a fair trial and several constitutional rights, including the right to be present, the right to counsel, the right to confront witnesses, the right to a public trial, and the right to a fair and impartial jury (Turner, 2021). At the same time, the introduction of virtual proceedings has also offered opportunities to improve court proceedings with the aid of technology (Rabinovich-Einy and Katsh, 2017).

With the wane of the pandemic, face-to-face proceedings have resumed, but it has become clear that online proceedings are here to stay, and debates over their appropriate scope have come to the fore, including specific concerns relating to criminal proceedings (Turner, 2021).

The rise of online courts

In the mid-2010s, a novel phenomenon emerged in several courts and tribunals in different parts of the world—traditional, face-to-face proceedings were being substituted with ones that relied on remote online communication. Interestingly, the principal motivation for the adoption of these processes was not making the courts more efficient—although the heavy backlog facing courts was certainly a consideration and an added benefit—but rather the promise for increased access to justice (Rabinovich-Einy and Katsh, 2017). To a large extent, the rise of online proceedings was centered on civil disputes. Indeed, ODR emerged in the e-commerce setting, as a specific solution for those disputes that arose online that were typically small scale in terms of value but could have significant impact on those individuals' confidence and trust in online mechanisms (Katsh and Rabinovich-Einy, 2017). As internet communication spread, technology improved, smartphones became widespread, and social networks proliferated, the distinction between online and offline disputes, and between what seemed appropriate to be handled online as opposed to face-to-face blurred (Rabinovich-Einy, 2021). These developments opened the door for addressing a broad range of disputes and proceedings online and, ultimately, for the adoption of ODR in court.

The early court ODR programs involved the adoption of fully redesigned court proceedings that took place asynchronously and in writing. In some processes, parties were referred to a 'diagnosis' stage—allowing them to better understand their rights and options through a series of tailored answers and questions, phrased in everyday simple language. In other cases, technology was used to offer tools for calculating amounts owed, drafting legal documents, or gathering relevant past decisions and out-

comes. Some of these processes were formal and engaged judges. Others were informal, involving online variations of negotiation and mediation, which employ algorithms to suggest texts to parties, highlight areas of agreement and disagreement, and to 'split the difference' or optimize their agreed-upon resolution (Katsh and Rabinovich-Einy, 2017). These processes emerged in the US, British Columbia, the Netherlands, England, China, and Singapore, to name a few pioneers (Rabinovich-Einy, 2021). In most of these places, ODR processes addressed small claims, traffic, and family disputes.

While generally targeting civil disputes, several exceptions stood out. In the US, a platform called Matterhorn (now known as Catalis), established by law professor J. J. Prescott, emerged from concerns over access to justice of disempowered parties with outstanding warrants. In this quasi-criminal context, it is the fear of showing up physically in court that drives such individuals to abstain from legal proceedings, resulting in serious consequences (Bulinsky and Prescott, 2016). Similarly, the traffic arena offers a non-civil context where access barriers, such as having to take time off work, have prevented individuals from contesting unjustified tickets, which can result in grave consequences, particularly for those parties belonging to disempowered groups (Crozier and Garrett, 2020).

During the second decade of the 21st century, online courts were expanding at a steady pace. In the US, their number was growing quickly, as two principal private companies—one operating across state courts while the other in county courts—could offer quick, professional, and scalable online solutions to courts looking to better manage their caseload and streamline the litigation pipeline (Rabinovich-Einy, 2021). Alongside the steady expansion of ODR in courts in the US and elsewhere, several concerns hampered the development of online courts in other places. Concerns ranged from abuse of justice and the digital divide (manifested in the lack of access to technology and/or the inability to effectively interact with online systems), to objections by the legal profession driven by concerns over unauthorized practice of law, fear of data security and privacy harms in the online environment (see Data Justice by Redden), and skepticism over the ability of online proceedings that rely on thin communication to fulfill our expectations from legal proceedings (Susskind 2019; Schmitz, 2020; Sela, 2018).

This state of affairs changed drastically in March 2020 with the onset of the COVID-19 pandemic. Many courts around the world were forced to shut down while others continued to operate under strict social distancing requirements and the reality of parties, attorneys, court personnel, and judges contracting COVID or being required to quarantine. To continue offering formal and informal dispute resolution processes, courts and other institutions quickly implemented 'remote proceedings' (Engstrom, 2021; Rabinovich-Einy, 2022). While some use was made of written and audio-based proceedings, the most prominent avenue for holding court proceedings in many places worldwide quickly became video (Susskind, 2019). In some places, all proceedings were handled remotely, while in others only urgent matters were addressed online.

Videoconferencing was of course not new to courts, criminal proceedings included. However, while in the past there were limited arenas into which video proceedings

were introduced, such as bail hearings, typically involving only the defendant connecting from afar, during COVID the scope of these proceedings grew dramatically. Remote proceedings expanded both in terms of the types of proceedings that employed video, and in terms of having all or several parties and, in some cases, the judge, participate remotely (Bannon and Keith, 2021).

Both ODR and remote proceedings fall under what can be termed 'online courts' in that such processes involve online communication and at least one of the parties communicates from afar. There are, however, important distinctions between these two avenues. While ODR harnesses technology to redesign legal processes around stakeholder needs, remote proceedings center on sustaining and making more efficient existing processes (Mentovich and Rabinovich-Einy, 2022; Rabinovich-Einy, 2022). As we show below, the goals one seeks to advance in the design and adoption of procedures are important and shape both parties' procedural experiences and substantive outcomes.

Access to justice, procedural fairness, legal outcomes, and legitimacy

While there are only a few empirical studies on the impact written asynchronous ODR proceedings have had on the process and party perceptions, the adoption of remote video-based court proceedings during COVID has resulted in a myriad of studies cited below that generated more knowledge on the impact of the shift to the online arena on various participants across different contexts. One important lesson learned is that context matters in choice of medium (face-to-face, video, or written ODR) for different types of parties (e. g., prisoners, tenants, teenagers) across different types of cases (criminal, civil, family). Context inevitably determines the impact medium will have on procedural justice, access, and legal outcomes, as well as different stages of the proceeding, and for different types of parties (Rabinovich-Einy, 2022).

While there is no study of written asynchronous ODR in the criminal context, there are several studies of traffic cases that study the impact of the shift online, which can provide valuable lessons. Several studies have established that introducing ODR into this context improved the efficiency and accessibility of courts (Prescott, 2017; Prescott and Sanchez, 2019). In addition, it was shown that remote traffic proceedings conducted through asynchronous written communication eliminated race-based and age-based outcome disparities relating to fine levels and charge reduction that existed in similar proceedings conducted in the physical courtroom (Mentovich et al., 2020). Other research established the significance of procedural justice perceptions, alongside perceptions of access to justice, in shaping perceptions of legitimacy in the online setting (Mentovich et al., 2023).

In addition to the study of written asynchronous ODR, research of video-conferencing proceedings explored the impact of the shift to synchronous video proceedings.

Pre-COVID, several studies uncovered findings demonstrating that defendants (or other parties) whose hearings were conducted remotely fared worse than similarly situated parties whose cases were held in court (Poulin, 2004; Walsh and Walsh, 2008; Diamond et al., 2010). Explanations ranged from increased bias towards those appearing on video, through difficulty of establishing credibility online, to opportunities missed due to the absence of a physical encounter (Bannon and Adelstein, 2020.

During COVID, studies and reports of remote video proceedings had mixed findings. While some hailed video proceedings for enhancing accessibility as participation rates of parties and witnesses increased (Ostrom et al., 2021), others emphasized the need to focus on the quality of participation. It was found that for participants that belong to disempowered groups and lack access to high-speed internet, are in prison, or live in small and crowded spaces, and typically access proceedings through phone rather than a computer, connecting to a remote proceeding and participating without interruptions proved challenging (Thornburg, 2020). Indeed, concerns arise regarding the ability to fulfill a sense of voice and whether participants feel that they have been treated with respect and in a neutral fashion. At the same time, one study drawing on mock video proceedings found that they could meet procedural justice standards, and even outperform the physical setting on certain parameters (Mulcahy et al., 2020).

The integration of technology into legal proceedings holds promise for various benefits but also gives rise to concerns over the realization of core rights, particularly within the realm of criminal proceedings. One prominent area of concern revolves around defendants' right to be present, often (but not universally) requiring consent for a defendant's remote participation (Turner, 2021). Another is the right to counsel and, relatedly, the right to confer with counsel privately. For represented parties, however, the new medium presented obstacles for consultation in confidence during proceedings and for those indigent parties who lack representation it became more difficult to access legal aid and obtain counsel (Benninger et al., 2021).

Another principal right that may be challenged by remote proceedings is the right of confrontation, raising many of the same concerns that the previous rights mentioned do. These rights encapsulate the longstanding notion that defendants possess the right to face their accusers in court, a concept that, in contemporary criminal justice, extends to the defendants' entitlement to confront those offering testimony against them (Benninger et al., 2021). Confrontation rights have historically operated within a formal, in-person setting where the jury, the witness, and the accused could mutually observe each other. The widespread shift to video in criminal trials raised concerns over the adequacy of confrontation through such medium for several reasons. For one, concerns arose over the ability of defendants (litigants in person and represented ones) to realize the right of confrontation given both access challenges and the thin nature of communication, as described above. Second, the notion that the formal and unmediated nature of face-to-face proceedings is more conducive to truthful testimony, casts doubt on the effectiveness of remote proceedings in uncovering untruthful testimony (Legg and Song, 2021). Finally, defendants whose trial was conducted from afar may feel that they receive an inferior opportunity to confront the charges.

This notion is significant because such experiences may harm their overall perception of the criminal justice system and willingness to comply with the law in the future (Sunshine and Tyler, 2003).

In terms of credibility of witnesses on video, two recent lab studies have found that the volume level of the speaker (Bild et al., 2021) and camera angles (Bannon and Adelstein, 2020 shape credibility assessments. These findings raise significant concerns vis-à-vis the right to a fair hearing (Turner, 2021), as speakers are often unaware of these elements and may not have control over them, certainly in the case of parties who are less versed in technology. Alongside concerns over the impact of the biases that could shape judicial impressions from the exposure to the living conditions of defendants and parties (Mentovich and Rabinovich-Einy, 2022; Rabinovich-Einy, 2022), other research has emphasized the equalizing effect that seeing defendants' living conditions over video could have in tearing the façade of equality that exists in the courtroom (Ferguson, 2022).

Finally, online proceedings have also stimulated debate on the realization of a public trial, both from defendants' and the public's perspective (Turner, 2021). Alongside concerns over loss over opportunities for participation, various jurisdictions have found different solutions ranging from restricting participation to particular members of the public or press representatives to live streaming the proceedings. In fact, some worry that the new medium may be too public, harming the privacy of defendants and others in much more invasive ways than physical hearings (Armstrong, 2020).

The above shows how a change of medium can dramatically affect access, perceptions of fairness, and outcomes of proceedings, but such impact may operate in different directions. While some of the findings conducted on asynchronous written ODR in traffic cases seemed quite positive in terms of access, procedural justice, and outcomes, the research on video proceedings (in various contexts) raises a more mixed picture, in particular for parties belonging to disempowered groups. Some research raises questions with respect to the accessibility, procedural fairness, and substantive fairness of video proceedings, while other research promises enhanced equality and better experiences. Interestingly, in terms of efficiency some empirical findings also question whether videoconferencing proceedings terminate more quickly than physical ones (Turner, 2021; Horton, 2022). Despite divergences in their evaluation, there is widespread agreement that remote proceedings are here to stay post-COVID, and in mediation and arbitration in the US and Canada, virtual proceedings are already becoming the norm (Schmitz, 2021).

Concluding remarks: future opportunities and challenges

Alongside the shift to virtual proceedings, another important technological change is taking place—the introduction of artificial intelligence into legal decision-making

(see Artificial Intelligence by Van Brakel). While these are two distinct developments, we can expect the two phenomena to reinforce one another. As more and more processes move online, the use of algorithms will grow (see Algorithm by Leese), initially in the procedural and decision-support realms, but more and more we can expect the line between support and recommendations on the one hand, and decision-making on the other hand, to fade.

Interestingly, the criminal arena has been one of the principal domains for the employment of algorithmic risk-assessment tools to generate recommendations for judges regarding a defendant's risk level and their likelihood of reoffending, in deciding whether to release or arrest them (Stevenson, 2018; see Sentencing and Risk Assessments by Ugwudike). These tools were a central reference point in the argument against the employment of artificial intelligence in legal decision-making. The recommendations made by these tools, the vast majority of which were adopted by judges, were found to be heavily biased (Re and Solow-Niederman, 2019; Mayson, 2019). Sources of bias in these systems stem, inter alia, from flawed training data fed into the algorithm, and various approaches have been put forth to address such deficiencies (e.g., Kroll et al., 2017). Some researchers, however, indicate that biases in this context run deeper and cannot easily be addressed (Crawford, 2021), while others suggest that relying on the black box of the human mind carries similar risks (Bonezzi et al., 2022).

As the fast-paced developments with ChatGPT have shown us, despite critiques and concerns, we can expect further developments in this domain. How these changes will shape the accessibility, procedural fairness, and the nature of outcomes produced in legal proceedings has yet to be seen. While the adoption of these mechanisms is already a reality, the particulars of their design and the context they will be adopted into can be premeditated and premised on rigorous research (Amsler et al., 2020).

Suggested reading

Katsh, M.E., & Rabinovich-Einy, O. (2017). *Digital Justice: Technology and the Internet of Disputes.* New York: Oxford University Press.

Mentovich, A., Prescott, J. J., & Rabinovich-Einy, O. (2023). Legitimacy and online proceedings: Procedural justice, access to justice, and the role of income. *Law & Society Review*, 57(2), 189 – 213.

Mulcahy, L. (2021). Virtual poverty? What happens when criminal trials go online? In D. Cowan & A. Mumford (eds.), *Pandemic Legalities* (pp. 41 – 52). Bristol: Bristol University Press.

Susskind, R. (2019). *Online Courts and the Future of Justice.* Oxford: Oxford University Press.

Turner, J. (2021). Remote criminal justice. *Texas Tech Law Review*, 53(2), 197 – 271.

References

Amsler, L. B., Martinez, J., & Smith, S.E. (2020) *Dispute System Design: Preventing, Managing, and Resolving Conflict.* Stanford, CA: Stanford University Press.

Armstrong, M. (2020). *Justice, Livestreamed.* [online] Slate. Available at: https://slate.com/technology/2020/08/zoom-courts-livestream-youtube.html.

Bannon, A., & Adelstein, J. (2020). *The Impact of Video Proceedings on Fairness and Access to Justice in Court.* Brennan Center for Justice

Bannon, A. L., & Keith, D. (2021). Remote court: Principles for virtual proceedings during the COVID-19 pandemic and beyond. *Northwestern University Law Review*, 115(6), 1875 – 1920.

Benninger, T., Colwell, C., Mukamal, D., & Plachinski. L. (2021). *Virtual Justice? A National Study Analyzing the Transition to Remote Criminal Court.* Stanford Criminal Justice Center.

Bild, E., Redman, A., Newman, E. J., Muir, B. R., Tait, D., & Schwarz, N. (2021). Sound and credibility in the virtual court: Low audio quality leads to less favorable evaluations of witnesses and lower weighting of evidence. *Law and Human Behavior*, 45(5), 481 – 495.

Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. *Journal of Experimental Psychology: General*, *151*(9), 2250 – 2258.

Bulinski, M., & Prescott, J. (2016). Online case resolution systems: Enhancing access, fairness, accuracy, and efficiency. *Michigan Journal of Race & Law*, 21(2), 205 – 249.

Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Xosts of Artificial Intelligence.* New Haven: Yale University Press.

Crozier, W. E., & Garrett, B. L. (2020). Driven to failure: An empirical analysis of driver's license suspension in North Carolina. *Duke Law Journal*, 69(7), 1585 – 1641.

Diamond, S. S., Bowman, L. E., Wong, M., & Patton, M. M. (2010). Efficiency and cost: The impact of videoconference hearings on bail decisions. *The Journal of Criminal Law & Criminology*, 100(3), 869 – 902.

Engstrom, D. F. (2021), Digital civil procedure. *University of Pennsylvania Law Review*, 169(8), 2243 – 2289.

Ferguson, A. G. (2022). Courts without court. *Vanderbilt Law Review*, 75(5), 1461 – 1522.

Horton, D. (2022). Forced remote arbitration. *Cornell Law Review*, 108(1), 204 – 137.

Katsh, E., & Rabinovich-Einy, O. (2017). *Digital Justice: Technology and the Internet of Disputes.* New York: Oxford University Press.

Katsh, E., & Rifkin, J. (2001). *Online Dispute Resolution: Resolving Conflicts in Cyberspace.* San Francisco: Jossey-Bass.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633 – 705.

Legg, M., & Song, A. (2021). The courts, the remote hearing and the pandemic: From action to reflection. *University of New South Wales Law Journal*, 44(1), 126 – 166.

Mayson, S. G. (2019). Bias in, bias out. *The Yale Law Journal*, 128(8), 2218 – 2300.

Mentovich, A., Prescott, J. J. and Rabinovich-Einy, O. (2023). Legitimacy and online proceedings: Procedural justice, access to justice, and the role of income. *Law & Society Review*, 57(2), 189 – 213.

Mentovich, A., & Rabinovich-Einy, O. (2022). *Diyunim Mishpatiyim Biy'mei Corona* [Legal proceedings in times of Covid]. *Mishpat U Mimshal*, 24(1), 283 – 305.

Mentovich, A., Prescott, J. J., & Rabinovich-Einy, O. (2020). Are litigation outcome disparities inevitable? Courts, technology, and the future of impartiality. *Alabama Law Review*, 71(4), 893 – 980.

Mulcahy, L., Rowden, E., & Teeder, W. (2020). Testing the case for a virtual courtroom with a physical jury hub: Second evaluation of a virtual trial pilot study conducted by JUSTICE. JUSTICE.

Ostrom, B., Douglas, J., Tallarico, S., & Shannon, R. (2021). *The Use of Remote Hearings in Texas State Courts: The Impact on Judicial Workload* (Final report). National Center for State Courts, Court Consulting Division.

Poulin, A. B. (2004). Criminal justice and videoconferencing technology: The remote defendant. *Tulane Law Review*, 78(4), 1089 – 1167.

Prescott, J. J. (2017). Improving access to justice in state courts with platform technology. *Vanderbilt Law Review*, 70(6), 1993 – 2050.

Prescott, J. J., & Sanchez, A. (2019). Platform procedure. In Y. Chang (ed.), *Selection and Decision in Judicial Process Around the World* (pp. 30 – 72). Cambridge: Cambridge University Press.

Rabinovich-Einy, O. (2021). The past, present, and future of online dispute resolution. *Current Legal Problems*, 74(1), 125 – 148.

Rabinovich-Einy, O. (2022). Process pluralism in the post-COVID dispute resolution landscape. *Texas Wesleyan Law Review*, 10(1), 55 – 73.

Rabinovich-Einy, O., & Katsh, E. (2017). The new new courts. *American University Law Review*, 67, 165 – 215.

Re, R. M., & Solow-Niederman, A. (2019). Developing artificially intelligent justice. *Stanford Technology Law Review*, 22(1), 242 – 289.

Schmitz, A. J. (2020). Measuring access to justice in the rush to digitize. *Fordham Law Review*, 88(6), 2381 – 2406.

Schmitz, A. J., (2021). Ordering online arbitration in the age of COVID … and beyond. *University of Missouri School of Law Legal Studies Research Paper* No. 2021 – 21.

Sela, A. (2018). Can computers be fair? How automated and human-powered online dispute resolution affect procedural justice in mediation and arbitration. *Ohio State Journal on Dispute Resolution*, 33(1), 91 – 148.

Sourdin, T., Li, B., & McNamara, D. M. (2020). Court innovations and access to justice in times of crisis. *Health Policy and Technology*, 9(4), 447 – 453.

Stevenson, M. T. (2018). Assessing risk assessment in action. *Minnesota Law Review*, 103, 303 – 384.

Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review*, 37(3), 513 – 548.

Susskind, R. E. (2019). *Online Courts and the Future of Justice.* Oxford: Oxford University Press.

Susskind, R. E., & Susskind, D. (2017). *The Future of the Professions: How Technology will Transform the Work of Human Experts.* Oxford and New York: Oxford University Press.

Thornburg, E. G. (2021). Observing online courts: Lessons from the pandemic. *Family Law Quarterly*, 54(3), 181 – 244.

Turner, J. (2021). Remote criminal justice. *Texas Tech Law Review*, 53(2), 197 – 271.

Walsh, F. M., & Walsh, E. M. (2008). Effective processing or assembly-line justice? The use of teleconferencing in asylum removal hearings. *Georgetown Immigration Law Journal*, 22(2), 259 – 284.

Nicholas Gibbs and Alexandra Hall

39 Online ethnography

Abstract: This chapter provides the reader with a guide to undertaking a criminological ethnography online. Drawing upon reflections from two social scientific projects, the work discusses the practicalities, ethical challenges and central debates that we experienced in the field. We draw the chapter to a close with some meditations on the future application of online ethnographic study and its potential as part of a multi-sited approach.

Keywords: online ethnography, qualitative research, digital ethics, lurking

Introduction

Ethnography is a well-established methodology in Criminology, offering qualitative researchers a periscope into diverse lifeworlds as they exist in time and place. With its roots in anthropology, the method offers a contextually rich and exploratory sensibility that is well attuned to the need of criminologists. Ethnographic work is valuable for those who are interested not only in the complexity and intimacy of social and cultural relations and rituals, but also the impact of wider socio-political institutions such as the law and state.

The foundation of ethnographic research is in-depth observation and interaction with participants, where the researcher is immersed in the subject's social world to achieve a truer understanding and appreciation of their lived experience, or what is often termed 'verstehen' (Ferrell, 1997). Traditional ethnographic field-sites are therefore found in homes, workplaces, or leisure settings (gyms, pubs, etc). However, reacting to the profound digitization of societies, criminological ethnographers have recently turned their gaze to online environments. The result is the influx of myriad novel ethnographic approaches that can be broadly characterized as 'online ethnography.' We define this development as traditional ethnographic methods modified to interact with online communities and environments (see Gibbs and Hall, 2021). This describes a researcher's concerted immersion in digital field-sites that can include online forums, social media sites, or darknet pages (see Darknet by Tzanetakis). Discussing the expansion of quantitative digital social science, Lavorgna and Myles (2022: 1579) argue that: "[i]n a context of 'digital positivism' based on big data research methods and computational criminology in online research, it is important to retain a space also for qualitative analyses focusing on interpretative and critical approaches." Online ethnography therefore represents a valuable means of redressing this balance and retaining the unique and adaptable qualitative insight offered by ethnographic study. Of course, not all ethnographic work necessarily needs to be critical, and online ethnography can also provide descriptive insight on the uses, politics, and practices of technology. What matters is that the ethnographic gaze is able to follow the flows of social life

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-040

and offer data that cannot be captured by alternative methodologies. In this chapter we will reflect on our own experiences to offer a guide to researchers carrying out online ethnographic work on issues of crime, harm, and control. First, counsel will be offered to those wishing to undertake an online ethnography, before some challenges and ethical considerations are laid out. We will end the chapter with some meditations on some innovations and future directions of the methodology, with a particular focus on its application as part of a multi-sited approach.

Conducting an online ethnography

It is first important to address the various terms associated with online ethnography. Delli Paoli and D'Auria (2021) note that online ethnography has been carried out under various guises, including 'ethnography on the Internet,' 'cyber ethnography,' and 'virtual ethnography.' Though subtly different in their approaches, each can be traced back to Kozinets' (1998) idea of netnography, which was devised to research consumer behavior on the internet. From these origins, online ethnography has been deployed to examine sites throughout the various stages of the internet, from forums and bulletin boards, to blogs, and social media platforms (Holt, 2020). We have therefore settled on the term 'online' here to broadly encapsulate ethnographic study undertaken in online spaces. What follows is informed by our own deployments of the method, which respectively examined the online market in illicit medicines (Hall and Antonopoulos, 2016) and the use and supply of image and performance enhancing drugs (Gibbs, 2023a).

Online ethnographic fieldwork can be split into two phases: passive and active. The 'passive' phase involves the researcher 'lurking' in the field, not actively shaping the research site and instead playing the role of a passive observer (Janetzko, 2008; Cera, 2023). Unsurprisingly, lurking is fraught with ethical challenges given that those being observed are unaware of the researcher's presence. Debate continues as to whether platforms like social media sites constitute public or private spaces (see Lingel, 2012) and therefore whether consent is required from those under study. If one assumes that digital spaces are no different to a public street, lurking can be understood as akin to a traditional ethnographer observing a culture at work. However, given that the researcher is hidden from the participant during the lurking process, they take on a decidedly more shadowy role. Though further ethical issues will be addressed later in this chapter, this contestation illustrates the numerous moral debates that circle online ethnography. Conversely, 'active' online ethnography describes the researcher becoming an active participant in the practices and culture of the online setting. This might be posting content, interacting with participants through direct messages, or liking and sharing online content.

In Hall and Antonopoulos' (2016) study, passive data collection took the form of lurking on platforms that facilitated the sale of illicit medicines, including online pharmacy surface web pages, forums and discussion boards, as well as social media sites.

This was initially facilitated by keyword searching that had been informed by a review of existing academic and grey literature, alongside insights garnered from experts. Similarly, Gibbs' (2023a) work, employed keyword searching on Facebook, Instagram, and surface web forums to locate and observe the mechanisms of the online market for image and performance enhancing drugs. This is consistent with other online ethnographies in criminology that have employed initial periods of lurking to observe online spaces without actively shaping the field (see Crowe and Hoskins, 2019; Addeo et al., 2021; Bakken et al., 2023).

We advocate for an initial lurking phase to familiarize oneself with the lexicon, norms, and conventions of an online community. Whether this be building familiarity with the terminology and use of emojis by a community of drug users or gaining insight into the nature of images posted by gang members on social media, peering into the spaces under study before leaping into the active phase holds a great deal of worth. Data for this passive aspect of online ethnography generally takes the form of screenshots and fieldnotes. We should note here that data obtained through passive online ethnography may be vulnerable to the researcher's assumptions or biases (see Bias by Oswald and Paul), given that while participants' voices are reported upon in some sense, not interacting with those actors leaves a considerable blind spot. For this reason, both studies that we lean on here employed active as well as passive phases.

The active phase of Hall and Antonopoulos' project saw the researchers create credible consumer profiles on social media and forums alongside email accounts. These served as the online 'face' presented to actors in the illicit medicine market and, given that the online ethnography was in part covert and supported by methods adopted across various field-sites, it allowed them to manufacture credible personas as prospective consumers to ascertain how illicit medicines were being marketed online. The ethics of this covert method will be discussed below. Gibbs, on the other hand, opted to use his genuine name and image when setting up profiles on Facebook and Instagram, which also showed his institutional affiliation and the purpose of his study (displayed in the profiles' 'bio' section). This candor reflected the researcher's need to keep consistency between his online and offline persona (given that the same participants were interacted with in both contexts) as well as his utilization of his own bodily and cultural capital as he posted 'training updates,' echoing the norms and practices under study. The decision to adopt this approach rather than simply operate a ghost account with no cultural coherency reflected Gibbs' position as a 'researcher participant' (Gans, 1967) immersed in a specific locale, which required a means of building authenticity. In such cases, the researcher's efforts are generally rewarded with enhanced access and participant trust (see Accessing Online Communities by Kaufmann). Relatedly, as noted by Bakken et al. (2023), digital capital is important in online drug markets and, while they discuss this specifically in relation to online sellers, we argue that a degree of digital literacy is just as important for research into consumer practices. Indeed, Gibbs noted that participants were reticent to engage with his profiles until they were suitably populated and trustworthy.

The digital environment, and specifically social media, also presents some opportunities for participant access and sampling that are lacking for traditional ethnographers. While the help of community gatekeepers remains important, Gibbs utilized the 'geotags'[1] from users' posts to identify participants in the local hardcore gym scene as well as searching for hashtags relating to certain gyms and brands of drugs. Again, it is worth addressing the ethical dimension of this method as, although users have made the decision to publicly link themselves to certain locations and brands, they did not do this with the understanding that these affiliations would qualify them as research participants. Gibbs therefore worked on the understanding that platforms like Instagram constituted public space and, given that these posts were accessible to anyone, understood the geotagging and hashtagging as public declarations worthy of study. Given how underdeveloped and emergent the field of digital ethics is (see Ethics by Markham), this is one of many examples in our work of situational ethics, wherein we advocate for a degree of flexibility under the mantra to 'do no harm' (see Gatson, 2011). Similarly, as both studies progressed, they utilized what Sidoti (2023) terms 'algorithmic sampling.' This is when the algorithms on social media sites are used to identify and follow accounts that are deemed similar, allowing us to connect with users and sellers alike. However, the in-built connectivity of social media sites also presented an ethical challenge with regard to participant confidentiality. This was played out in Gibbs' study when a childhood friend stumbled upon his researcher Instagram account:

> I logged onto my research Instagram account to be greeted by a notification telling me that my friend, Tommy, had somehow found and started to follow me. Clearly this is a concern as, due to my sample hopefully following me back soon, he could easily find out who my participants are if he wanted to. When setting up the account I was careful to untick the options to link the profile to my phone number or personal email address, so I am unsure as to how he found it. I sent him a message and he duly unfollowed me, apologising and saying that the account had come up as a recommended follow. (Fieldnote, 26/06/19)

As demonstrated above, online ethnographers must remain vigilant of such exposure if they choose to present as their authentic identities online. Though this infraction was ultimately resolved, Tommy might have compromised numerous participants' confidentiality.

Following the sampling process, both studies approached online participants via the specific platform's messaging service to either pose as an interested customer or overtly interview the user/supplier about their experiences (see Recruitment via Social Media by Andersen). Given that both studies employed elements of covert ethnographic practice, some ethical justification is required. Working under a framework of situational ethics, we deemed the use of some deceptive practice necessary to glean the

1 Users on Instagram can choose to 'tag' their photos to certain geographic areas like cities, specific streets, or even individual premises. For Gibbs' study, he examined posts that geotagged in the gyms studied as part of the offline element of the research.

data on illicit market interactions that would be out of reach otherwise. Indeed, Gibbs' initial interactions with image and performance enhancing drug suppliers—within which he revealed his identity as a researcher—proved unfruitful as sellers did not engage. Instead, posing as an interested customer led to far more valuable data which ultimately informed the project's assessment of the market. Crucially, following Nissenbaum's (2010) situational ethics framework of 'contextual integrity,' which describes a 'case-by-case' ethical assessment, he found that making initial enquiries under the guise of a prospective customer posed little ethical risk. These interactions were limited to questions about the available products and pricing and did not move beyond initial enquiries (see Ferguson, 2017 for more on this). Similarly, all data were anonymized, and pseudonyms were attributed to the sellers to avoid any risk of exposure.

Ultimately, data collected in the reactive phase of both studies was made up of online interview transcripts, screenshots of various interactions, images and text posted by participants, documents that were hosted on the platforms under study, and fieldnotes. In line with the covert aspects of both projects, some of these data were collected without explicit consent of those under study (for example, transcripts of conversations between the researchers posing as customers as well as posts that were in the public realm). Further, in line with the messy nature of ethnographic study, active and passive phases inevitably bled into one another as the researcher seeks to capture the essence of the digital landscape.

Challenges and ethical considerations

Previous work has identified various ethical and practical issues with online ethnography (see Gibbs and Hall, 2021; Thompson et al., 2021) but, for the sake of brevity we will focus here on *building trust, informed consent*, and *data searchability.* Gibbs (2023b), in an article stemming from the project described above, emphasizes the need for sellers to build trust with their customers in online markets. The underlying reason for this is the scope for deception that exists online, particularly if one is studying illicit digital markets. Illicit actors are often drawn towards the cloak of anonymity that the online environment provides, particularly in disparate commodity-focused markets like those for fake medicines or image and performance enhancing drugs.

The task for the online ethnographer, therefore, is to approach such participants who will be distrusting and may even assume that the researcher is undercover law enforcement. This is compounded by the challenge of not being 'seen' in the same sense that traditional ethnographers were when embedded in the physical spaces of the community. Although Gibbs somewhat addressed this with his authentic profiles and content generation, many participants still chose not to respond, presumably because they feared the exposure of their involvement in illicit activities. Similarly, though the affordances of platforms like social media offer ease of access to researchers, ethical questions about non-culturally embedded ethnographers descending upon an online community ought to be posed. Unlike offline spaces, where the ethnographer

must often earn the trust of participants and engage in norms and customs, the lack of barriers to digital spaces opens them up to researchers who are outsiders. Do we lose some valuable ethnographic tradition if such researchers can simply parachute into an online community and take but ultimately do not give back? Equally, even if a researcher lacks 'insider' status, if they have engaged with literature and enter the field in an open-minded manner, is this really much of a departure from traditional ethnographic practice? A further consideration is the reliability of anonymous online participants' accounts (Hine, 2008). Although some participants might feel emboldened to speak more honestly due to their concealed identity (Paech, 2009), researchers must be critical given that they cannot be sure of the participant's authenticity.

Secondly, as hinted at earlier in this chapter, the challenge of ensuring that informed consent is obtained and navigating the boundaries between public and private data is key. Particularly when researching on social media platforms, it is often unclear whether data such as statuses and posts are public, given their searchability, or private, given that the user has presumably made the content just for their friends and contacts (Morey et al., 2012; Cera, 2023). Scholars writing on digital ethics have tended to either advocate for explicit informed consent to be given regardless of the nature of the data (boyd, 2008), or jettisoned the need for this in line with a view that, when anonymized, what is posted online constitutes public property (Kitchin, 2003). This debate speaks to wider issues around participants' awareness and ownership of their online footprint and personal data in an age of surveillance capitalism (Zuboff, 2015). This contestation fundamentally raises a few questions; given that users produce data for various purposes—for example, sharing information or views with friends and followers—why does the researcher have the right to use this material as 'data' in the first place? And, when producing this data, is the user aware that they may unwittingly become a 'participant'? With this said, it is well-documented that participants' behavior changes when they are aware that they are being observed, so are covert approaches in fact a necessity if we are to capture a true picture of the social world? With regard to privacy, both studies we reflect on here encountered numerous 'semi-private' groups on Facebook and drug forums. After much consideration, we opted to work with these data on the condition that it was fully anonymized.

Finally, the issue of data searchability is worthy of consideration. In the world of Web 2.0, what happens on the web really does *stay* on the web. Therefore, 'trace' data (Geiger and Ribes, 2011) is often searchable and, while this can be a helpful means of online ethnographic insight, it also represents an ethical dilemma. As data typically includes verbatim quotations, platforms like forums, non-private social media accounts, and surface web pages are accessible to the public, meaning that one need only search for the quotation to unmask the participant's identity. Given that participants received a guarantee of confidentiality in both studies, a mitigation strategy was needed. Such a strategy is Markham's (2012) notion of 'fabrication as an ethical practice,' whereby searchable direct quotations are superficially altered so as to prevent any ethical breaches. This was particularly important for Gibbs as he examined promotional hash-

tags employed by sellers which, if disseminated verbatim, would have jeopardized a multitude of actors involved in the illicit market.

Ultimately, as Gatson (2011) contends, online ethnographers cannot rely on a one-size-fits-all approach to ethics and instead should engage in something akin to situational ethics by trusting themselves to follow the mantra of 'do not harm' (Barbosa and Milan, 2019). In this sense, both projects applied a malleable ethical framework based more on participant protection and harm minimization than the 'little other' safeguards (Winlow and Hall, 2011) typical of contemporary academia. Above all else, we wish to emphasize the nascent nature of digital ethics and would encourage budding online ethnographers to remain pragmatic and risk-aware.

Innovations, applications and future directions

Even in its infancy, we have seen a slew of innovations in the practice of online ethnography. Foremost has been an understanding that online ethnographic study ought not to be divorced from offline practices and can instead sit comfortably alongside more 'traditional' methods (Walker, 2010). Aiming to quash the online/offline dualism often present in criminological research, we advocate for an eclipse of the barriers between digital and physical spaces and for online sites to be viewed as part of the wider tapestry of one's journey through the world of 'ubiquitous computing' (Duggan, 2017). In this sense, rather than reinforcing the dualistic conception of online versus offline spaces, we suggest that online ethnography holds great potential for carrying out what Marcus (1995) terms multi-sited ethnographic study. This is supported by Mackenzie (2022: 1538) who, discussing the 'metaverse' and the future of digital/physical space, argues:

> [E]ventually, the line between the real and the simulated will no longer be a particularly relevant or helpful distinction to make for analysts of the social. While we currently talk of online advancements like 'virtual reality' or 'augmented reality', the future is said to be more likely to be 'mixed reality': a mix between on and offline.

Looking forward, we might be better served discussing online ethnography on 'metafields' (Airoldi, 2018) as the nature of the digital field-site itself is set to become more contested. Encouragingly, in their systematic review of online ethnographic work, Delli Paoli and D'Auria (2021) found that one third of the research adopted such a bricolage approach, employing interviews, offline ethnography, and techniques like focus groups to complement the method. An awareness of the place of online ethnography in relation to offline spaces somewhat futureproofs the method as well as highlighting its utility in keeping pace with an ever-changing social context. Importantly, this multi-sited approach does not eclipse the need to study online space or devalue the contribution of research undertaken wholly in online environments. Instead, it highlights the value of understanding the productive tensions and kinship of online and offline space.

Online ethnography has been utilized to explore a range of disparate criminological phenomena in recent years, including financial crime (Mackenzie, 2022), incel culture (Thorburn et al., 2023), alcohol consumption (Romo-Avilés et al., 2023), and even the relationship between real estate and criminality (Maoz and Gutman, 2024). The terrain is now ripe for students and researchers of criminology to add to this expanding body of work by going beyond the Web 2.0 spaces of social media and digital commerce, and instead turning a critical eye to the criminogenic potential of artificial intelligence and other emergent themes (Hayward and Maas, 2020).

Conclusion

This chapter has served as a guide for anyone wishing to carry out a criminological online ethnography. The method acts as a counterbalance to the swell of 'big data' studies and ensures that criminology can retain an immersive qualitative focus amid a period of techno-social proliferation. We have offered some of the promises and pitfalls of this approach, from improved access to hard-to-reach communities through to the challenges of informed consent. Ultimately, as long as criminological researchers retain a commitment to situational ethics and an exploratory ethnographic sensibility, this method can offer unique insights into diverse criminogenic phenomena in a rapidly changing social world.

Suggested reading

Abidin, C., & de Seta, G. (2020). Private messages from the field: Confessions on digital ethnography and its discomforts. *Journal of Digital Social Research*, 2(1), 1 – 19. 10.33621/jdsr.v2i1.35.

Delli Paoli, A., & D'Auria, V. (2021). Digital ethnography: A systematic literature review. *Italian Sociological Review*, 11, 243 – 267. https://doi.org/10.13136/isr.v11i4s.434.

Gibbs, N., & Hall, A. (2021). Digital ethnography in cybercrime research: Some notes from the virtual field'. In A. Lavorgna & T. J. Holt (eds.), *Researching Cybercrimes* (pp. 283 – 299). Cham: Palgrave Macmillan.

Sidoti, C. (2023). Smart researching in criminology: Virtual ethnography at the edge. In R. Faria & M. Dodge (eds.), *Qualitative Research in Criminology* (pp. 53 – 67). Cham: Springer. https://doi.org/10.1007/978-3-031-18401-7_4

References

Addeo, F., Punziano, G., & Padricelli, G. (2021). Using digital methods to shed light on 'border phenomena': A digital ethnography of dark tourism practices in time of COVID-19. *Italian Sociological Review*, 11(4S), 269 – 291.

Airoldi, M. (2018). Ethnography and the digital fields of social media. *International Journal of Social Research Methodology*, 21(6), 661 – 673. https://doi.org/10.1080/13645579.2018.1465622

Bakken, S. A., Oksanen, A., & Demant, J. (2023). Capital in illegal online drug markets: How digital capital changes the cultural environment of drug dealing. *Theoretical Criminology*, 27(3), 421 – 438. https://doi.org/10.1177/13624806221143365

Barbosa, S., & Milan, S. (2019). Do not harm in private chat apps: Ethical issues for research on and with WhatsApp. *Westminster Papers in Communication and Culture*, 14(1), 49 – 65. https://doi.org/10.16997/wpcc.313

boyd, d. (2008). Facebook's privacy trainwreck. *Convergence: The International Journal of Research into New Media Technologies*, 14(1), 13 – 20. https://doi.org/10.1177/1354856507084416

Cera, M. (2023). Digital ethnography: Ethics through the case of QAnon. *Frontiers in Sociology*, 8. https://doi.org/10.3389/fsoc.2023.1119531.

Crowe, N., & Hoskins, K. (2019). Researching transgression: Ana as a youth subculture in the age of digital ethnography. *Societies*, 9(3), 53. https://doi.org/10.3390/soc9030053.

Delli Paoli, A. and D'Auria, V. (2021). Digital Ethnography: A systematic literature review. *Italian Sociological Review*, 11, 243 – 267. https://doi.org/10.13136/isr.v11i4s.434

Duggan, M. (2017). Questioning 'digital ethnography' in an era of ubiquitous computing. *Geography Compass*, 11(5), e12313. https://doi.org/10.1111/gec3.12313

Ferguson, R.-H. (2017). Offline 'stranger' and online lurker: Methods for an ethnography of illicit transactions on the darknet. *Qualitative Research*, 17(6), 683 – 698. https://doi.org/10.1177/1468794117718894

Ferrell, J. (1997). Criminological verstehen: Inside the immediacy of crime. *Justice Quarterly*, 14(1), 3 – 23. https://doi.org/10.1080/07418829700093201.

Gans, H. (1967). *The Levittowners: Ways of Life and Politics in a New Suburban Community.* New York: Columbia University Press.

Gatson, N. (2011). The methods, politics, and ethics of representation in online ethnography. In N. Denzin, & Y. Lincoln (eds.), *The SAGE Handbook of Qualitative Research*, 4th ed. (pp. 513 – 527). Thousand Oaks, CA: SAGE Publications Ltd.

Geiger, R., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. In *2014 47th Hawaii International Conference on System Science*, pp. 1 – 10. Washington, DC: IEEE. 10.1109/HICSS.2011.455.

Gibbs, N., & Hall, A. (2021). Digital ethnography in cybercrime research: Some notes from the virtual field. In A. Lavorgna & T. J. Holt (eds.), *Researching Cybercrimes* (pp. 283 – 299). Cham: Palgrave Macmillan.

Gibbs, N. (2023a). *The Muscle Trade: The Use and Supply of Image and Performance Enhancing Drugs.* Bristol: Bristol University Press.

Gibbs, N. (2023b). #Sponseredathlete: The marketing of image and performance enhancing drugs on Facebook and Instagram. *Trends in Organized Crime.* https://doi.org/10.1007/s12117-023-09491-4

Hall, A., & Antonopoulos, G. A. (2016). *Fake Meds Online: The Internet and the Transnational Market in Illicit Pharmaceuticals.* London: Palgrave Macmillan.

Hayward, K., & Maas, M. (2020). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, 17(2), 209 – 233.

Hine, C. (2008). Virtual ethnography: Modes, varieties, affordances. In N. Fielding, R. Lee, & G. Blank (eds.), *The SAGE Handbook of Online Research Methods* (pp. 257 – 270). London: SAGE.

Holt, T. (2020). Assessing traditional and nontraditional data collection methods to study online phenomena. In C. Melde & F. Weerman (eds.), *Gangs in the Era of Internet and Social Media* (pp. 1 – 19). Cham: Springer.

Janetzko, D. (2008). Nonreactive data collection on the internet. In N. Fielding, R. M. Lee, & G. Blank (eds.), *Online Research Methods* (pp. 161 – 174). London: Sage.

Kitchin, H. A. (2003). The Tri-Council Policy Statement and research in cyberspace: Research ethics, the internet, and revising a 'living document'. *Journal of Academic Ethics*, 1(4), 397 – 418. https://doi.org/10.1023/b:jaet.0000025671.83557.fa

Kozinets, R. (1998). On netnography: Initial reflections on consumer research investigations of cyberculture. *Advances in Consumer Research*, 25(1), 366 – 371.

Lavorgna, A., & Myles, H. (2022). Science denial and medical misinformation in pandemic times: A psycho-criminological analysis. *European Journal of Criminology*, 19(6), 1574 – 1594. doi:https://doi.org/10.1177/1477370820988832.

Lingel, J. (2012). Ethics and dilemmas of online ethnography. In CHI'12 *Extended Abstracts on Human Factors in Computing Systems*, pp. 41 – 50.

Mackenzie, S. (2022). Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial. *The British Journal of Criminology*, 62(6), 1537 – 1552. https://doi.org/10.1093/bjc/azab118

Maoz, E., & Gutman, M. A. (2024). The capitalization of crime in the city of real estate. *Theoretical Criminology*, 28(1), 28 – 49. https://doi.org/10.1177/13624806231161595

Marcus, G.E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology*, 24(1), 95 – 117.

Markham, A. (2012). Fabrication as ethical practice. Qualitative inquiry in ambiguous Internet contexts. *Information, Communication & Society*, 15(3), 334 – 353. https://doi.org/10.1080/1369118X.2011.641993

Morey, Y., Bengry-Howell, A., & Griffin, C. (2012). Public profiles, private parties: Digital ethnography, ethics and research in the context of Web 2.0. In S. Heath & C. Walker (eds.), *Innovations in Youth Research* (pp. 195 – 209). London: Palgrave Macmillan.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford: Stanford Law Books.

Paech, V. (2009). A method for the times: A meditation on virtual ethnography faults and fortitudes. *Nebula*, 6(4), 195 – 215. http://www.nobleworld.biz/images/Paech.pdf

Romo-Avilés, N., Pavón-Benítez, L., & Tarancón Gómez, P. (2023). 'Keeping your composure': A digital ethnography of gendered alcohol norms on Instagram. *International Journal of Drug Policy*, 112, 103936. https://doi.org/10.1016/j.drugpo.2022.103936

Sidoti, C. (2023). Smart researching in criminology: Virtual ethnography at the edge. In R. Faria & M. Dodge (eds.), *Qualitative Research in Criminology* (pp. 53 – 67). Cham: Springer. https://doi.org/10.1007/978-3-031-18401-7_4.

Thompson, A., Stringfellow, L., Maclean, M., & Nazzal, A. (2020). Ethical considerations and challenges for using digital ethnography to research vulnerable populations. *Journal of Business Research*, 124, 676 – 683. https://doi.org/10.1016/j.jbusres.2020.02.025.

Thorburn, J., Powell, A., & Chambers, A. (2023). A world alone: Masculinities, humiliation and aggrieved entitlement on an incel forum. *The British Journal of Criminology*, 63(1), 238 – 254. https://doi.org/10.1093/bjc/azac020

Walker, D.M. (2010). The location of digital ethnography. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 2(3), 23 – 39. https://doi.org/10.5130/ccs.v2i3.1596

Winlow, S., & Hall, S. (2011). What is an 'Ethics Committee'?: Academic governance in an epoch of belief and incredulity. *British Journal of Criminology*, 52(2), 400 – 416. https://doi.org/10.1093/bjc/azr082

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75 – 89. https://doi.org/10.1057/jit.2015.5

Simon Egbert

40 Platforms

Abstract: This chapter focuses on one of the latest technical developments (not only) in police work: the emergence of data integration and analysis platforms. The police use these platforms to seamlessly combine data from different sources and to make it easier to analyze. These platforms are updating the myth that lots of data promises relevant findings and are also reproducing an association-centered logic of knowledge production. This is closely linked to the risk of increased surveillance.

Keywords: data integration and analysis platforms, Palantir, police, network, surveillance, algorithms

This chapter focuses on a technical tool that plays an increasingly important role in digital societies: data integration and analysis platforms. Most, if not all, criminological relevant organizations in digitized societies will (have to) work with such or similar tools in the future. Because these platforms will be used to cope with the constantly increasing mass of data and to maintain analysis capability. Therefore, this chapter provides an overview of the current proliferation of data integration and analysis platforms with special emphasis on platforms from the US software company Palantir and their utilization in the field of policing.

While the phenomena related to the term 'digital platform' are very diverse, they still share some basic traits: a connectivity-oriented infrastructure that aims to facilitate interactions by at least two third parties and a mode of functioning based on massive and diverse data, as well as a modular architecture (e. g., Andersson Schwarz, 2017; van Dijck et al., 2018). With this definition in mind, data integration and analysis platforms are understood here as digital platforms that aim at making available heterogenous data from different sources, especially by facilitating cross-connections and desilozation to generate relevant and actionable associations.

While focusing on platformized policing in this chapter with regard to data integration and analysis platforms, it needs to be highlighted that already other understandings on platformized policing exist (see Policing by Wilson). For example, Wil-

Disclaimers and Acknowledgments: This chapter has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 833749). I acknowledge support for the publication costs by the Open Access Publication Fund of Bielefeld University and the Deutsche Forschungsgemeinschaft (DFG).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-041

son's (2019, 2021) work on the increasing use of cloud services by police forces or Gates' (2019) analysis on the role of Axon as a police platform, trying to generate revenue with (police) data, ultimately seeking to redesign the police organization as a platform in its entirety (see also Wood, 2019). It must also be emphasized that not only the police, but also other criminologically relevant organizations, especially intelligence services (see Intelligence by Gundhus and Lundgaard), use these technologies (Biddle, 2017). And not only that, but such platforms are also already being used in other, quite different areas of society, for example by the United Nations World Food Programme (Matak, 2021) or the National Health Service (NHS) in the United Kingdom (Mason, 2023). Although the US-software company Palantir, which can be considered the current market leader of data integration and analysis platforms for policing with its platform named 'Gotham,' is the main reference throughout the text, other comparable solutions do exist, like the open-source 'KNIME' Analytics Platform (KNIME, 2024), Hitachi Vantara's 'Pentaho' Data Integration and Analytics platform (Hitachi Vantara, 2024), Atos' 'research and analysis platform' (Atos, 2022), or 'rsShadow' by Rola security solutions (Rola Security Solutions, 2024).

The argumentation of this chapter is structured as follows: first, we will present a short overview of the use of platforms in police agencies of different countries around the world, which leads us to the thesis that police are becoming increasingly platformized. Second, we will focus on the key functionalities of data integration and analysis platforms. After that, we will discuss the societal risks of police forces using these platforms. In the conclusion, we summarize our argument and provide pointers for future digital-criminological research.

The platformization of policing: an empirical overview

As is so often the case when it comes to new technological developments in policing, US police forces are the pioneers of platformization. The first known Police Departments using Palantir Software come from the Salt Lake Valley, in which Palantir software was used *inter alia* by the Salt Lake City Police Department since 2010 (Palantir, 2012). Especially referring to the Palantir Graph application, the police force focused on investigating criminal networks, using data from different sources: mug shots, airport data, criminal records, suspicious activity reports, among others (Hardy, 2016; Palantir, 2012). Also, the Los Angeles Police Department (LAPD) was early adopting Palantir's Gotham platform (Brayne, 2021; Ferguson, 2017). The LAPD used it from 2011 onwards, when it played a major role in the Los Angeles' Strategic Extraction and Restoration Program, also known as Operation LASER, which aimed at reducing gun-related violence by focusing on gang members and repeat violent offenders (Ferguson, 2017: 102–103; Uchida et al., 2012). The platform was used to create so called 'Chronic Offender Bulletins,' overviews of the criminal careers of relevant offenders, plus to produce information

on their spatial and personal environment, accessing police data like information from license plate readers or field information cards (Brayne, 2017: 986–987). In 2012, Palantir software was implemented by the local police in New Orleans, as part of the 'NOLA for Life' initiative launched by Mayor Mitchell J. Landrieu, which aimed at implementing a "comprehensive murder reduction strategy" (City of New Orleans, 2013). Palantir's Gotham software was used in this for building a unified data environment consisting of internal police data banks and public data sources, most notably police and public safety data as well as information on infrastructure, like the location of liquor stores, hospitals, or parks (Ferguson, 2017: 41).

In other countries, data integration and analysis platforms are implemented and used by police forces as well. In Norway, for example, Palantir's Gotham platform was purchased in 2016 to connect data from 19 different police databanks in a project with the characteristically name 'Omnia' (meaning 'everything')—which has cost close to 100 million Norwegian kroner (approx. US$14 million) but has ultimately failed (Gundhus and Wathne, 2024). Palantir Gotham is also used by the Danish police, where the system is called POL-INTEL, becoming fully operational in 2018 (Galis and Karlsson, 2024). In Germany, three federal police forces use modified versions of Palantir's Gotham (Ulbricht and Egbert, 2024). Palantir software has also been used at the supranational police level, namely by Europol from 2016 to 2021 (European Parliament, 2020).

Key characteristics of data integration and analysis platforms

What specifically makes data integration and analysis software so attractive for police forces and what technical characteristics are associated with this?

The primary goal of these platforms is to make analysis and research processes faster by connecting heterogenous data from different sources and by making them accessible from one central digital location. Being a special type of digital platform (see Ulbrich and Egbert, 2024), data integration and analysis platforms share important similarities with the well-known internet-based digital platforms such as Amazon, Facebook, or Uber (van Dijck et al., 2018). However, they also have important differences: instead of connecting third parties and receiving fees for this service, as in the case of for example Uber or Airbnb, data integration and analysis platforms connect their clients with data, as well as data (banks) among each other. By doing so, they try to enhance interoperability, the "so-called smart way to connect the dots and to avoid continuing to work in 'silos' with segmented information networks" (Bigo, 2020: 410), which is reflected in Palantir's main goal of the 'desilozation' of information systems (Palantir, n.d.: 2).

Of course, desilozation is not a value in itself, but follows a specific rationality: by being able to look at different data sources at the same time and to relate the data they

contain with each other, new knowledge shall be produced. In the field of police, this idea has been known for years and has already emerged along with the first algorithmic analysis procedures (see e. g., McCue and Parker, 2003) and is updated by Palantir by offering their platforms "as the connective tissue between an organization's data, its analytics capabilities, and operational execution" (Palantir, 2020).

An illustrative example of the knowledge creation potential connected to data integration and analysis platforms is the already mentioned graph application from Palantir's Gotham platform. It provides a space for performing link analysis and is one of the most critical applications for communicating (provisional) analytical conclusions. The graph application implements the link analysis and association-centered analytical approach of Palantir Gotham in a very direct way, it can be thus characterized as its analytical core. Via a spider diagram architecture, it allows to show associations between people or objects in a visually convenient and actionable way, allowing to identify relevant connections by police analysts at first glance (Brayne, 2021).

However, data integration and analysis platforms are not only about producing new knowledge, but also about speeding up decision-making processes and rendering police work more efficient. Providing a unified data environment, data integration and analysis platforms make heterogenous and numerous data sources available with one click, instead of enabling access only individually, with recourse to separate programs and formal access procedures (Iliadis and Acker, 2022).

And even if there are many points of criticism associated with these platforms (see below), the criticism of automation that is otherwise often expressed with regard to algorithmic decision-making systems is only partly applicable here (e. g., Algorithm Watch, 2020). This is the case because data integration and analysis platforms imply a special socio-technical interaction, meaning that algorithms and analysts *together* generate new knowledge and, hence, inform decisions. And although the platforms themselves play an important role in these processes, similar to their better-known platform siblings from the internet (like YouTube and Facebook), data integration and analysis platforms are commonly framed as neutral intermediaries (e. g., Gillespie, 2010). In the words of Palantir: "We build digital infrastructure for data-driven operations and decision-making. [...] We license this software to organisations, who receive secure and unique instances of our platforms in which to conduct their own work on their own data" (Palantir, 2020). As opposed to dismissing the platforms' generative and transformative role, we should, however, understand these platforms as mediators, as van Dijck (2013: 29) suggests.

Risks

A risk which is connected to policing in general, but to platformized policing in specific, refers to surveillance (see Surveillance by Lyon). As described, the goal of data integration and analysis platforms is the connection of diverse data sources to make existing data more easily analyzable by police officers. Surveillance is the immediate flip side of

this approach, because the operational idea of data integration and analysis platforms is built on the well-known big data-myth that one only needs enough data to discover new, hitherto unknown connections (e. g., Anderson, 2008). This approach implies that every attempt should be made to utilize and integrate as much data as possible into the platforms. This can ultimately serve as incentive to interpret legal access restrictions to data as narrow as possible, and, if necessary, also to exceed these limits (Egbert, 2019). This is one of the reasons why Iliadis and Acker (2022) name Palantir's software a 'surveillance platform.'

One practical implication of the close relationship of these platforms and surveillance is what Brayne (2021: 112) calls "net widening," meaning that the number of persons who come into the police's focus will increase, since police analysts will increasingly use platforms to check persons who have not yet been processed by the police and who have so far only been indirectly involved with the police, for example witnesses or associates of serial offenders (friends, family, etc.). This is based on an approach which Brayne (2021: 39) terms "dragnet surveillance," that is the "collection and analysis of information on everyone, rather than only people under suspicion," which implies lowering the threshold above which citizens are subject to police control. Although Brayne refers this approach to digital policing in general, it applies to data integration and analysis platforms in specific, because they provoke the reversion of an important principle of police work with regard to civil rights: Instead of checking a person more closely when a suspicion against them arises in the course of the investigation, people are now monitored in order to establish such a suspicion in the first place.

A special problem associated with data integration platforms given the low barrier to entry into police data systems can be called 'risk/suspicion by association.' This principle means that an individual is perceived as suspicious or risky because—according to the data available to the platform—it interacts with (past) offenders and/or known high-risk persons. Since information about the social milieu and entourage play an important role in the social network- or "associational suspicion"-approach facilitated by data integration and analysis platforms, their utilization not only indirectly harbors the risk to materialize this principle (Ferguson, 2017: 200).

According to Brayne (2021: 116), dragnet surveillance also includes that those people who were already in the police's focus before, are now being monitored even more closely, since their data are analyzed across increasing and diverse institutions (besides police, also social services, or banks, for example). This in turn can lead to a 'vortex' of disadvantage, because people who are more intensely controlled by the police have increasingly worse chances of, for example, getting a loan (Brayne, 2021; Eubanks, 2017). And this dynamic is also a problem precisely because police data tends to be systematically biased, since police officers in many countries of the world tend to be racially biased in their control and reporting behaviour, which then inevitably affects the quality of the police database and the decisions that are made based on this data (e. g., Richardson et al., 2019). In fact, the LASER project of LAPD was criticized precisely because of the 'vicious circle' it creates with the help of Palantir platforms by

integrating low-threshold data (e. g., field interview cards) that are collected based on the (biased) spatial and person-related selection decisions made by police officers (Bhuiyan, 2021). The use of platforms, then, ultimately leads to precisely these areas and people coming under even more police scrutiny, since platforms necessarily identify those areas or people as more suspicious that are overrepresented in the data sets accessible to them.

The socio-technical constellation connected to data integration and analysis platforms mentioned above entails that it is particularly difficult to reconstruct its knowledge creation and decision-making processes retrospectively, which is, ultimately, detrimental to the transparency of police work. However, although the human part in this process is important, we should not make the mistake of devaluing the platforms' share of the analytical work. Even the compilation and processing of data represents an analytical intervention that should not be underestimated in its epistemic relevance (Galis and Karlsson, 2024). Data integration and analysis platforms specify which data can be considered by human analysts and, hence, with what probability they are processed. In addition, the very representation of connections is already charged with suspicions and risk attributions, with at least subliminally reproducing a logic of 'risk/suspicion by association.'

A final risk that is associated with data integration platforms is connected to the question of what happens to those data or results that have been viewed and generated with the help of the platform. A dispute between the New York Police Department (NYPD) and Palantir in 2017 shows that it can lead to serious problems when such collaborations end. In this specific case, the NYPD decided to use a new software and they needed analytical data from their past utilization of Palantir software from the company in a format which was readable by their new software, which Palantir repeatedly refused to deliver, invoking intellectual property (Alden, 2017). Although it is unknown how this dispute ended, it nonetheless shows the potential problems associated with such platforms as they intervene deeply into the everyday police work such as knowledge creation and decision-making. As a current example of the dangers of the 'privatization of policing' through technologization, platformization shows that the use of technology by the police can always be accompanied by a loss of control and autonomy, which ultimately represents a risk for both the police and the citizens (Byrne and Marx, 2011: 31; see Privatization by Lomell).

Conclusion

The main take away points of this chapter are:

– Due to the increasing volume of data that organizations in general and police authorities in specific are exposed to, data integration and analysis platforms will become important tools in the day-to-day knowledge creation activities of police officers.

- By connecting heterogeneous data sources and dismantling boundaries between databases (desilozation), data integration and analysis platforms primarily aim to enhance interoperability.
- Data integration and analysis platforms intend to support and facilitate analytical actions by police officers, e. g., by showing associations between people in a visually convenient and actionable way.
- The approach of data integration and analysis platforms by definition is closely connected to issues of surveillance, since they aim to make as much data as possible accessible.
- Platformization provokes the reversion of an important civil rights principle: instead of checking a person more closely on the basis of concrete suspicion, people are now inspected via platforms to establish such a suspicion in the first place.
- Data integration and analysis platforms demonstrate once again the dangers of the privatization of policing through technologization, because they penetrate deep into the everyday work of police officers. This is ultimately linked to handing over a significant part of the knowledge generation to technologies and private companies and, thus, to the risk of losing autonomy and control.

Suggested reading

Brayne, S. (2021). *Predict and Surveil. Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Ferguson, A. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement.* New York: New York University Press.

Galis, V., & Karlsson, B. (2024). A world of Palantir – ontological politics in the Danish police's POL-INTEL. Information, Communication & Society, 1 – 19. https://doi.org/10.1080/1369118X.2024.2410255

Gundhus, H. O. I., & Wathne, C. T. (2024). Resistance to platformization: Palantir in the Norwegian police. *Information, Communication & Society*, online first, 1 – 19. https://doi.org/10.1080/1369118X.2024.2325533

Iliadis, A., & Acker, A. (2021). The seer and the seen: Surveying Palantir's surveillance platform. *The Information Society*, 38(5), 334 – 363.

Ulbricht, L., & Egbert, S. (2024). In Palantir we trust? Regulation of data analysis platforms in public security. *Big Data & Society*, 11(3), 1 – 15. https://doi.org/10.1177/20539517241255108

References

Alden, W. (2017, June 28). *There's a fight brewing between the NYPD and Silicon Valley.* BuzzFeed News. https://www.buzzfeednews.com/article/williamalden/theres-a-fight-brewing-between-the-nypd-and-silicon-valley

Algorithm Watch. (2020). *Automating Society Report 2020.* https://automatingsociety.algorithmwatch.org

Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired.* https://www.wired.com/2008/06/pb-theory/

Andersson Schwarz, J. (2017). Platform logic: An interdisciplinary approach to the platform-based economy. *Policy & Internet*, 9(4), 374 – 394. https://doi.org/10.1002/poi3.159

Atos. (2022). *Aufbau einer souveränen Recherche- und Analyseplattform für die Polizeien.* https://atos.net/wp-content/uploads/2022/09/WP-Aufbau-einer-souveraenen-Recherche-und-Analyseplattform-fuer-die-Polizeien.pdf

Bhuiyan, J. (2021). LAPD ended predictive policing programs amid public outcry. A new effort shares many of their flaws. *The Guardian.* https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform

Biddle, S. (2017). *How Peter Thiel's Palantir Helped the NSA Spy on the Whole World.* The Intercept. https://theintercept.com/2017/02/22/how-peter-thiels-palantir-helped-the-nsa-spy-on-the-whole-world/

Bigo, D. (2020). Interoperability: A political technology for the datafication of the field of EU internal security? In D. Bigo, T. Diez, E. Fanoulis, B. Rosamond, & Y. A. Stivachtis (eds.), *The Routledge Handbook of Critical European Studies* (pp. 400 – 417). London: Routledge. https://doi.org/10.4324/9780429491306

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977 – 1008. https://doi.org/10.1177/0003122417725865

Brayne, S. (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Byrne, J., & Marx, G. (2011). Technological innovations in crime prevention and policing. A review of the research on implementation and impact. *Cahiers Politistudies*, 3(20), 17 – 40.

City of New Orleans. (2013). *NOLA for Life. Comprehensive Murder Reduction Strategy.* https://www.nola.gov/nola/media/Mayor-s-Office/Files/NOLA%20FOR%20LIFE/NOLA-FOR-LIFE-Sept2013-book-pdf_1.pdf

Egbert, S. (2019). Predictive policing and the platformization of police work. *Surveillance & Society*, 17(1/2), 83 – 88. https://doi.org/10.24908/ss.v17i1/2.12920

Eubanks, V. (2017). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York: St. Martin's Press.

European Parliament. (2020). *Answer given by Ms Johansson on behalf of the European Commission (9 June 2020).* https://www.europarl.europa.eu/RegData/questions/reponses_qe/2020/000173/P9_RE(2020)000173_EN.pdf

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement.* New York: New York University Press.

Galis, V., & Karlsson, B. (2024). A world of Palantir – ontological politics in the Danish police's POL-INTEL. Information, Communication & Society, 1 – 19. https://doi.org/10.1080/1369118X.2024.2410255

Gates, K. (2019). Policing as digital platform. *Surveillance & Society*, 17(1/2), 63 – 68. https://doi.org/10.24908/ss.v17i1/2.12940

Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, 12(3), 347 – 364. https://doi.org/10.1177/1461444809342738

Gundhus, H. O. I., & Wathne, C. T. (2024). Resistance to platformization: Palantir in the Norwegian police. *Information, Communication & Society*, online first, 1 – 19. https://doi.org/10.1080/1369118X.2024.2325533

Hardy, Q. (2016). The risk to civil liberties of fighting crime with big data. *The New York Times.* https://www.nytimes.com/2016/11/07/technology/the-risk-to-civil-liberties-of-fighting-crime-with-big-data.html

Hitachi Vantara. (2024). *Pentaho Data Integration and Analytics. Ingest, Integrate and Analyze.* https://www.hitachivantara.com/en-us/products/pentaho-plus-platform/data-integration-analytics.html

Iliadis, A., & Acker, A. (2022). The seer and the seen: Surveying Palantir's surveillance platform. *The Information Society*, 38(5), 334 – 363. https://doi.org/10.1080/01972243.2022.2100851

KNIME. (2024). *KNIME Analytics Platform.* https://www.knime.com/knime-analytics-platform

Mason, R. (2023). Everything you need to know about NHS England's biggest ever IT contract. *The Guardian.* https://www.theguardian.com/society/2023/oct/12/everything-you-need-to-know-about-nhs-englands-biggest-ever-it-contract

Matak, V. (2021, August 2). *WFP and Palantir partner to tackle Covid-19 and world hunger* [Video]. World Food Programme (WFP). https://www.wfp.org/stories/video-wfp-and-palantir-partner-tackle-covid-19-and-world-hunger

McCue, C., & Parker, A. (2003). Connecting the dots: Data mining and predictive analytics in law enforcement and intelligence analysis. *The Police Chief*, 70(10), 115 – 122.

Palantir. (n.d.). *NOLA Murder Reduction. White Paper.* https://s3.documentcloud.org/documents/21032274/nola-murder-reduction-white-paper.pdf

Palantir. (2012). *Fighting Crime in Salt Lake City.* www.palantir.com/wp-assets/wp-content/uploads/2012/06/ImpactStudy_SLCPD.pdf

Palantir. (2020, November 12). Palantir is not a data company (Palantir Explained, #1) [Medium]. *Palantir Blog.* https://medium.com/palantir/palantir-is-not-a-data-company-palantir-explained-1-a6fcf8b3e4cb

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYU Law Review.* https://www.nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/

Rola Security Solutions. (2024). *rsShadow. Shines a light in the darkness.* https://www.rola.com/en/rsshadow/

Uchida, C. D., Swatt, M., Gamero, D., Lope, J., Salazar, E., King, E., Maxey, R., Ong, N., Wagner, D., & White, M. D. (2012). *The Los Angeles Smart Policing Initiative: Reducing Gun-Related Violence through Operation LASER.* Los Angeles Police Department. https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/losangelesspi.pdf

Ulbricht, L., & Egbert, S. (2024). In Palantir we trust? Regulation of data analysis platforms in public security. *Big Data & Society*, 11(3), 1 – 15. https://doi.org/10.1177/20539517241255108

van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media.* Oxford: Oxford University Press.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The Platform Society: Public Values in a Connective World.* Oxford: Oxford University Press.

Wilson, D. (2019). Platform policing and the real-time cop. *Surveillance & Society*, 17(1/2), 69 – 75. https://doi.org/10.24908/ss.v17i1/2.12958

Wilson, D. (2021). The new platform policing. In A. Završnik & V. Badalič (eds.), *Automating Crime Prevention, Surveillance, and Military Operations* (pp. 47 – 68). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-73276-9_3

Wood, S. E. (2019). Policing through platform. *Computational Culture*, 7 (21 October 2019). http://computationalculture.net/policing-through-platform/

Dean Wilson

41 Policing

Abstract: Police digitalization has the capacity to produce significant transformations in policing practice and organization. This chapter examines the potential impacts on police work and on those policed of digital technologies, suggesting possible directions for future research.

Keywords: digitalization, predictive policing, platformization, police technologies

The digitalization of policing is one of the most significant trends in contemporary criminal justice. Prior strategies such as community policing, problem-oriented policing, hot-spot policing, and intelligence-based policing all encouraged the leveraging of data and analysis to inform police operations. Nevertheless, contemporary police digitalization exemplified through strategies such as 'predictive policing' and 'precision policing' does herald novel transformations and challenges (see Prediction by Ķīlis, Gundhus and Galis). The volumes and combinations of datasets, advances in predictive analytics and data mining, and competitive markets in security technology, have all congealed to mobilize an intensive datafication of police work that continues to accelerate. Predictive policing—which uses predictive analytics to allocate and guide patrols—has received most scholarly attention to date. Nevertheless, predictive policing software is now often only one component of larger cloud-based information architectures that seek to integrate public and private databases, and a range of police technologies including Body Worn Cameras, sensorized weapons and uniforms, ANPR data, and GPS tracking. Sometimes referred to as 'Big Data Policing' this movement towards complex and mutable public-private socio-technical assemblages is also captured in recent discussions of 'platform policing' and 'platformization' (Wilson, 2019a, 2019b, 2021; Egbert, 2019; Gates, 2019; Linder, 2019; see Platforms by Egbert). Undergirded by elements of path dependency and ideological 'dataism' (Van Dijck, 2014) holding that predictive accuracy increases through the compilation, accumulation, and processing of ever larger and more varied datasets, processes of police digitalization exhibit an expansionary logic.

This entry will first examine the specific drivers of contemporary police digitalization. It will then examine the principal areas of research and debate in the field, examining the potential for discrimination and bias, the obscuring of police decision-making, the reception and deployment of technologies in specific contexts, and the organizational implications of police digitalization. This entry draws primarily—though not exclusively—on studies of predictive policing where most research attention has focused to date but will conclude by suggesting possible agendas for ongoing research into police digitalization. While policing online spaces might also be considered an aspect of police digitalization, it is not covered in this chapter.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-042

The rise of police digitalization

While contemporary police digitalization is novel in its scale and analytic capacity, the aspiration that technology could transform police agencies into efficient organizations equipped for the challenges of policing contemporary societies is certainly not. Police cars and two-way radios in the 1930s, computerized Command and Control Systems in the 1960s, and organizational computerization from the 1960s were all accompanied by vivid imaginaries of efficient and modernized policing that founded for various social, political, and technological reasons (Wilson, 2019b). While these earlier entanglements of policing with technology form part of the genealogy of police digitalization, there are several historical contexts from earlier this century that have informed the emergence and rapid advance of police digitalization. The attacks upon the World Trade Centre in 2001 had profound impacts across the security field that cascaded into operational policing. 9/11 energized a pre-emptive orientation, a drive towards data collection and database integration, intensified surveillance, an expansion of the market for security technologies and the formation of 'surveillant assemblages' linking disparate public and private actors and data. Another key event underpinning police digitalization was the Global Financial Crisis of 2008. The collapse of the global financial system resulted in a fiscal crisis and a decade of fiscal austerity where public services—including police agencies—endured sharply reduced budgets. The then novel innovation of predictive policing offered a solution. While science-fiction inspired allusions to seeing into the future were peppered throughout trade literature, it was the promise of enabling police departments to do 'more with less' (Wilson, 2018) that was equally alluring and became an important factor in the rise of predictive policing. Predictive policing embodied the zeitgeist of austerity with its vision of lithe cost-efficient police agencies delivering 'just-ahead-of-time' policing (Wilson, 2018).

The evolution of predictive policing into platform policing emerged in the United States, where the widespread adoption of Body Worn Cameras following the shooting of Michael Brown in Ferguson, Missouri in 2014 created significant storage and processing issues through the volume of visual data. Cloud storage providers—including Tech goliaths Amazon and Microsoft—offered cloud-based public safety storage and processing platforms where police agencies could purchase a range of technologies through 'software-as-a-service' models (Linder, 2019). Most evident in the United States, but unfolding across policing agencies globally, digitalization is advanced by technology vendors and police leaders as a panacea for the plethora of problems besetting policing in the contemporary context—police violence and bias, declining public trust, low workforce morale, and stretched budgets (Ferguson, 2017; Wilson, 2019b). This has energized a powerful 'socio-technical imaginary' (Jasanoff, 2015) of police digitalization that envisages digital technology driving reinvented policing that is proactive, efficient, precise, transparent, and responsive.

Critiques of police digitalization

These new conceptions of policing as data-driven, predictive, and reliant upon algorithmic calculations have received considerable scrutiny from critical scholars. There are three main critiques that are intertwined: 1) a tendency towards what Andrejevic (2018) terms 'frameless data collection'—a faith that predictions and calculations should be continually refined by accumulating ever larger and more diverse datasets, regardless of their obvious relationship to crime or law enforcement; 2) the potential to entrench and amplify existing inequalities and discrimination in policing, through digitally inscribing flawed practices and statistics via data feedback loops; and, 3) the obscuring and rendering unaccountable of police decision-making which remains concealed within inscrutable algorithmic calculations that are claimed to be neutral and objective (see Algorithm by Leese). Concerns that 'frameless data collection' would ensue initially rested in predictive policing's evident affinity with the commercial sector's entrancement with Big Data (see Big Data by Završnik), and its corresponding adulation of data volume, the integration of data sets, data mining, predictive analytics, and the favoring of correlation over causation (Wilson, 2018). It is a valid critique, and the inherent amplification of police surveillance capacity it represents is manifest in empirical contexts. While early predictive policing software such as PredPol© used only a limited number of data points, expansive surveillant assemblages such as the NYPD Domain Awareness System partnership with Microsoft integrate vast and disparate data streams which can then be subjected to machine analysis (Ferguson, 2017). Processes of platformization extending from predictive policing software have fueled a significant data thirst, as police agencies endeavor to integrate disparate data sources in a quest for infinite searchability and automatically generated insights (Egbert, 2019).

The second critique argues that police digitalization and platformization will both entrench and intensify extant patterns of discrimination, criminalization, and stigmatization in policing. This critique extends from the observation that police statistics are notoriously incomplete and error-prone and, moreover, to a large extent reflect police activity and discretion (Wilson, 2018; Bennett Moses and Chan, 2018). Police statistics emerge from specific social processes that are in themselves frequently problematic. The implication of this is that allocating police patrols based upon past police data to predict future incidents will trigger escalating spirals of police activity and data collection, as automated allocations direct patrols to geographic areas that have been the focus of intensive police presence in the past (Ferguson, 2017; Brayne, 2021; Wilson, 2018). While some software developers have attempted to address this problem through random patrol allocations (Shapiro, 2019) this remains a trenchant critique of police digitalization. While critical analysis has most frequently addressed place-based deployments in predictive policing, person-based predictive techniques have proved equally—if not more—controversial. Many initial experiments with person-based predictive applications, such as Chicago's Strategic Subject List, Los Angeles LASER, and the New Orleans Palantir system, have been discontinued due to civil rights concerns,

ineffectiveness, and maintenance issues (Marciniak, 2023). Nevertheless, demonstrating a propensity to extend data collection out through social networks—enacting what Brayne (2021) terms 'secondary surveillance'—such person-based digital databases have been critiqued for energizing racialized and marginalizing police tactics (Ferguson, 2017; Jefferson, 2020). While the body of empirical research on person-based databases remains limited (for examples see Marciniak, 2023; Brayne, 2021), the focus of recently touted strategies such as precision policing on person-based tactics enabled through digital technology (Haberman et al., 2022) render this an important field for future research.

The third critique, closely related to the others, is the obscuring and masking of police decision-making and guiding rationales through reference to machine calculations that are presented as impartial and objective (Bennett Moses and Chan, 2018; Wilson, 2018; Kaufmann et al., 2019). Empirical research does indicate that data-driven police operational decisions risk becoming increasingly opaque as proprietary algorithms and machinic calculations stretched across socio-technical assemblages both mystify and decenter decision-making processes (Kaufmann et al., 2019). Such opacity is also intensified through the interpenetration of private commercial interests with public policing agencies (Brayne, 2021). Although presenting a considerable challenge, acknowledging the three intertwined problems attendant with police digitalization, numerous scholars have advanced pragmatic policy recommendations for police organizations. Both Ferguson (2017) and Brayne (2021) outline the potential of police data, if made publicly available, to facilitate greater transparency and accountability in policing, particularly in relation to police stops and police use of force. Moreover Ferguson (2017), Egbert and Leese (2021), and Van Brakel (2021) all provide useful guidelines that include oversight, transparency, and risk assessment on community impact that may curtail the more deleterious possibilities and actualities that can accompany police digitalization.

Digitalizing the police organization: contexts and contestation

While the three critiques focus on police-community impacts of police digitalization, research also focuses on the impacts and transformations within police organizations. Technology vendors, drawing freely from science-fiction imagery, often promote visions of a seamless field of interwoven technologies. These digital assemblages, it is often claimed, will not only lead to utopias of public safety, but also produce nimble and flexible policing agencies with enhanced officer safety, less paperwork, and more support for officers in the field (Wilson 2019a, 2021). Nevertheless, earlier research into police and information and communication technologies already indicated that police digitalization in practice would not resemble the public safety utopias or dystopias evoked by critics and boosters alike. Dampening initial enthusiasm for the

promised efficiencies of information and communication technologies, initial ethnographic studies revealed that impacts remained highly contingent upon local conditions and organizational structures. The overall shaping of individual socio-technical systems was mediated by the selection of technology, the focus of individual police units and the expertise and abilities available within police organizations (Chan, 2003: 658). While some research suggested a negligible impact upon operational policing (Manning, 2008), others suggested significant—though ambiguous—effects, including the flattening of hierarchies, information overload, the over-collection of data, and resistance to and subversion of technology by police in the field (Chan, 2003).

Recent ethnographic research into police digitalization suggests both continuity with these earlier cautions, and the need to be wary of flamboyant claims emanating from the security technology sector. Sufficient research evidence into police digitalization has now amassed to allow more nuanced and empirically informed assessments. It is perhaps predictive policing that has received the most research attention to date. One of the aspirations for predictive policing—and for digitalization generally—is that machine calculations can replace the flawed and biased judgments of individual officers, an attractive proposition for those calling for police reform (Shapiro, 2019), given that many contemporary problems in policing have been attributed to police discretion and decision-making. Research data reveals however that human decision-making remains significant in the face of digitalization (Fussey et al., 2021). Sandhu and Fussey's (2021) research examining the deployment of predictive policing software in a British police force found considerable resistance from officers who questioned the utility and accuracy of the digital calculations and remained suspicious of the potential of the technology to the subjective skills that constituted policing 'craft.' Ratcliffe et al.'s study of HunchLab predictive software in Philadelphia reported similar suspicion amongst some officers who subscribed to more traditional approaches to patrol. Nevertheless, their research revealed varied reactions to the software, with some officers perceiving it as a useful extension and aid to police work (Ratcliffe et al., 2020). Applying insights drawn from Science and Technology Studies, researchers have noted that digitalization occurs within specific social and organizational contexts. Digital technology is transforming police work. Nevertheless, these transformations are mediated through co-constitutive processes as the technology is engaged within complex socio-technical assemblages where actors contest, interpret and repurpose data informed by diverse values and objectives (Fussey et al., 2021).

Resistance to and lack of engagement with digital technologies from officers in the field is a persistent theme in the literature. Technologies may be resisted by officers if they are perceived to be flawed or not useful (Sandhu and Fussey, 2021). Another significant reason why digital technologies may be resisted and subverted however is the fear of automation (see Automation by Mann). Officer concerns that policing is being 'deskilled' as decision-making and discretion are transferred to automated processes resurfaces repeatedly in the research (Egbert and Leese, 2021; Sandhu and Fussey, 2021), and some of the more exuberant socio-technical imaginaries of police digitalization suggest it is a not entirely unfounded concern (Wilson, 2019a).

Resistance and subversion of digitalization from the police workforce may also result from it being perceived as a form of managerial surveillance. The introduction of new technologies into policing has historically been accompanied by new modes of monitoring and surveilling officers (Wilson, 2019b). While such potential remains implicit in the trade literature, others have not been as reticent to expound the management potential of police digitalization. Bratton and Malinowski's vision of COMPSTAT Plus, drawing upon the labor management practices of major retailers such as Target Corporation, imagines sensorized environments constantly processing raw data "in real time according to a series of integrated performance metrics that measure and then display weighted data" (2008, p. 264). While algorithmically modulating workflow to increase officer productivity (Benbouzid, 2019), predictive policing software and other technologies augment the capacity to monitor officers' spatial and temporal locations, with the digital assemblages of 'platform policing' often incorporating visual dashboards integrating and displaying data from sensors, cameras, and other technologies recording officer activity (Wilson, 2021). Unsurprisingly perhaps, Brayne's (2021) ethnography of the LAPD revealed subversion and resistance to digital managerial surveillance from officers in the field. Nevertheless, police digitalization may produce deeper structural transformations that foreshadow tighter managerial control and a narrowing of discretion. Examining the digitalization of policing in Norway extending from intelligence-led strategies, Gundhus et al. (2022) present evidence that the standardization and specialization attendant with data-driven policing can result in intensified managerial centralization and control through 'digital taylorism' and a concomitant diminution of feelings of autonomy and expertise amongst front-line officers.

Conclusion

Police digitalization continues apace. It is therefore imperative that criminologists develop research agendas and conceptual tools to understand the new policing landscape as it unfolds. Importantly, police digitalization involves the formation of complex socio-technical assemblages that weave together dynamic and sometimes internally fractious constellations of public and private actors, databases, software, and hardware. To conclude, there are five significant dimensions of police digitalization that might inform ongoing criminological critique and research. These are:

- Broad claims that police digitalization will produce organizational efficiencies, procedural fairness, community engagement, and enhanced public safety should be subject to both conceptual scrutiny and empirical evaluation from criminologists and empirically examined and critiqued.
- There is vital work still to be undertaken in critically assessing the capacity of police digitalization to inscribe, obscure, intensify, and create patterns and practices of discrimination and inequality in specific milieus.
- It is evident police digitalization has the capacity to produce significant transformations in police organizations and in police practice, energizing novel socio-tech-

nical assemblages, and producing novel forms of expertise, assessment, contestation, and negotiation. Criminologists can fruitfully engage with cognate fields such as science and technology studies, surveillance studies, and organization studies to examine the network dynamics and operational impacts of specific sites and technologies.
- Police digitalization is a global phenomenon, and research into Global South contexts (see for example Duarte, 2021; Narayan, 2023) is of considerable importance.
- Researchers should also continue to examine the political economy of police digitalization, an aspect captured by the concept of 'platform policing.' Invaluable work has been undertaken on the role of technology designers, developers, and vendors in police digitalization (Shapiro, 2019; Linder, 2019), and trade shows and literature also remain invaluable sources.

Police digitalization is propelled by vivid socio-technical imaginaries. The contours of these imaginaries, and how they are fractured and reinvented when colliding with the empirical realities of policing, are fertile areas for criminological inquiry.

Suggested reading

Brayne, S. (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford: Oxford University Press

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London: Routledge.

Fussey, P., Davies, B., & Innes, M. (2021). 'Assisted' facial recognition and reinvention of suspicion and discretion in digital policing. *British Journal of Criminology*, 61(2), 325 – 344.

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *British Journal of Criminology*, 59(3), 674 – 692.

Wilson, D. (2019a). Platform policing and the real-time cop. *Surveillance & Society*, 17(1/2), 69 – 75.

References

Andrejevic, M. (2018). Data collection without limits: Automated policing and the politics of framelessness. In A. Završnik (ed.), *Big Data, Crime and Social Control* (pp. 93 – 107). London: Routledge.

Bennett Moses, L., & Chan, J., (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*, 28(7), 806 – 822.

Benbouzid, B. (2019). To predict and to manage. Predictive policing in the United States. *Big Data & Society*, 6(1), 1 – 13.

Bratton, W., & Malinowski, S. (2008). Police performance management in practice: Taking COMPSTAT to the next level. *Policing*, 2/3, 259 – 265.

Brayne, S. (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press

Chan, J. (2003). Police and new technologies. In T. Newburn (ed.), *The Handbook of Policing* (pp. 655 – 679). Cullompton: Willan.

Duarte, D. E. (2021). The making of crime predictions: Sociotechnical assemblages and the controversies of governing future crime. *Surveillance & Society*, 19(2), 199 – 215.

Egbert, S. (2019). Predictive policing and the platformization of police work. *Surveillance & Society*, 17(1/2), 83 – 88.

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London: Routledge.

Ferguson, A. (2017). *The Rise of Big Data Policing: Surveillance, Race and the Future of Law Enforcement.* New York: NYU Press.

Fussey, P., Davies, B., & Innes, M. (2021). 'Assisted' facial recognition and reinvention of suspicion and discretion in digital policing. *British Journal of Criminology*, 61(2), 325 – 344.

Gates, K. (2019). Policing as digital platform. *Surveillance & Society*, 17(1/2), 63 – 68.

Gundhus, H., Talberg, N., & Wathne, C. (2022). From discretion to standardization: Digitalization of the police organization. *International Journal of Police Science & Management*, 24(1), 27 – 41.

Haberman, C., Bratton, W., Murad, J., & Rawlins, W. (2022). *Precision Policing 2.0: A Framework for the Future of Policing.* Cincinnati, OH: University of Cincinnati Institute of Crime Science.

Jasanoff, S. (2015). Future imperfect: Science, technology and the imaginations of modernity. In S. Jasanoff & K. Sang-Hyun (eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (pp. 1 – 33). Chicago: University of Chicago Press.

Jefferson, B. (2020). *Digitize and Punish: Racial Criminalization in the Digital Age.* Minneapolis, MN: University of Minnesota Press.

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *British Journal of Criminology*, 59(3), 674 – 692.

Linder, T. (2019). Surveillance capitalism and platform policing: The surveillance assemblage-as-service. *Surveillance & Society*, 17(1/2), 76 – 82.

Manning, P. K. (2008). *Crime Mapping, Information Technology and the Rationality of Crime Control.* New York: New York University Press.

Marciniak, D. (2023). Algorithmic policing: An exploratory study of the algorithmically mediated construction of individual risk in a UK police force. *Policing & Society*, 33(4), 449 – 463.

Narayan, S. (2023). *Predictive Policing and the Construction of the 'Criminal': An Ethnographic Study of Delhi Police.* London: Palgrave Macmillan.

Ratcliffe, J., Taylor, R., & Fisher, R. (2020). Conflicts and congruencies between predictive policing and the patrol officer's craft. *Policing and Society*, 30(6), 639 – 655.

Sandhu, A., & Fussey, P. (2021). The 'uberization of policing'? How police negotiate and operationalise predictive policing technology. *Policing & Society*, 31(1), 66 – 81.

Shapiro, A. (2019). Predictive policing for reform? Indeterminacy and intervention in big data policing. *Surveillance & Society*, 17(3/4), 456 – 472.

Van Brakel, R. (2021). Rethinking predictive policing: Towards a holistic framework of democratic algorithmic surveillance. In M. Schuilenberg & R. Peeters (eds.), *Algorithmic Societies: Power, Knowledge and Technology in the Age of Algorithms* (pp. 104 – 118). London: Routledge.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197 – 208

Wilson, D. (2018). Algorithmic policing: The futures of predictive policing. In A. Završnik (ed.), *Big Data, Crime and Social Control* (pp. 108 – 127). London: Routledge.

Wilson, D. (2019a). Platform policing and the real-time cop. *Surveillance & Society*, 17(1/2), 69 – 75.

Wilson, D. (2019b). Predictive policing management: A brief history of patrol automation. *New Formations: A Journal of Culture/Theory/Politics*, 98, 139 – 155.

Wilson, D. (2021). The new platform policing. In A. Završnik & V. Badalič (eds.), *Automating Crime Prevention, Surveillance and Military Operations* (pp. 47 – 68). Dordrecht: Springer.

Emils Kilis, Helene O. I. Gundhus, and Vasilis Galis

42 Prediction

Abstract: Prediction has a long history in the social sciences, and advances in computing and statistics have transformed our ability to predict in a wide range of domains. However, concerns have been raised about an indiscriminate application of a predictive logic, and crime is an area where this is quite pronounced. Indeed, while the police, correctional service, and criminal courts have become increasingly reliant on digital systems of prediction, critics have drawn our attention to numerous issues and complexities attendant to this process. This chapter looks at prediction in the criminological realm and provides an overview of key arguments concerning the way data are generated, organized, and used as input for predictive tools and technologies, and how the results are interpreted in the context of criminal justice. By doing so, it aims to show that the discussions surrounding prediction highlight how digital tools are transforming the nature of knowledge and expertise within the criminal justice system.

Keywords: prediction, predictive policing, criminal justice system, risk assessment

Introduction

Prediction involves the use of data to make claims and inferences about future events or outcomes, and it has become an integral aspect of various domains of social life. This need to predict was also evident at the birth of the modern social sciences and sovereign state. Thinkers, social activists, and public officials sought to deploy the new disciplines for the purposes of managing societies and rectifying social ills (including crime), based on insights derived from social data, particularly demographic data. This was underpinned by the philosophical belief that social phenomena were, in principle, amenable to statistical analysis, prediction and, ultimately, control (see Porter and Ross, 2008). For instance, Michel Foucault (2003) notes the emergence of biopolitics at the end of the 18th century and the attendant desire to measure various social phenomena in statistical terms. The population became a predictable and manageable political problem, and complex issues related to statecraft became legible and simplified.

The ability of social science to provide useful and meaningful predictions has long been a contested topic. Digitalization, however, has revitalized the promise of prediction (Aradau and Blanke, 2018). The advent of sophisticated computational tools has

Disclaimers and acknowledgments: The research has been funded under NordForsk Research and Innovation Programme on Digitalisation of the Public Sector and is an output of the project No. 100786: Critical Understanding of Predictive Policing (CUPP). The project investigated how institutional and social values, digital affordances, and organizational politics are conceived and embedded in data-driven police innovations. From 2020 to 2024, the project explored how these are experienced and practiced by police officers and developers of digital police infrastructure in Denmark, Estonia, Latvia, Norway, Sweden, and the UK.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-043

opened new possibilities for practitioners in various fields, including the social sciences, to gather and analyze data, and interpret social reality (Halpern, 2015). This has led to renewed optimism about the capacity to predict and forecast different phenomena, including crime and criminal behavior (Hardyns and Rummens, 2017). Predictions of worst-case scenarios and catastrophic consequences provides a moral platform for politics, security, and criminal justice personnel to act with decisiveness (McCulloch and Wilson, 2016). Relying on predictions assumes that risks can be calculated and that risk populations can be identified according to objective criteria, based on reliable data (Zedner, 2009). However, contemporary pre-emptive strategies assume hypothetical future crime in a way that differs from calculable predictions. Consequently, there is a need to study the phenomenon of prediction empirically and critically (see Bennett Moses and Chan, 2018; Brayne, 2017; Kaufmann et al., 2021) and avoid speculative technopositivism.

To accommodate these different positions on prediction in criminology and broader academic discussions, we should understand prediction in context, focusing on the object of prediction, the way these predictions are produced and their limitations. Hence, we argue that a full appreciation of the nature of prediction requires us to approach predictive technologies as boundary objects.

Prediction and digitalization

Recent advances in computing and statistics have transformed our ability to predict in a wide range of domains. These changes have been greeted with both enthusiasm and criticism and should be understood as part of a broader shift towards preventive and anticipatory logics. Digitalization has been said to enable the collection, storage, and analysis of vast amounts of data. Predictive analytical tools have allowed practitioners in various fields to harness this data and identify patterns, trends, and correlations that would likely be impossible to discern via human observation. This, in turn, enhances our ability to predict future outcomes based on historical data and knowledge of the field in question, and develop preventive and pre-emptive measures.

These transformations have impacted different spheres including business (Brynjolfsson et al., 2019), justice (Oster, 2021), insurance (Cevolini and Esposito, 2020), healthcare (Rehman et al., 2022), public administration (Kersing et al., 2022), and migration (Scheel et al., 2019). Predictive models based on financial data and economic indicators shape decisions about asset allocation and risk management, and identify anomalies that may indicate, for example, fraudulent activities. In health care, predictive tools enable health care professionals to identify individuals at risk of developing certain diseases, while insurance underwriting models can be used to calculate insurance premiums. Likewise, digital solutions enable public health officials to monitor and analyze population health data to identify trends, develop targeted interventions, and allocate resources.

However, there have also been a plethora of critical approaches that highlight the implications of predictive and preventive logics in a variety of sensitive contexts, such as security and justice. It has been argued that algorithmic management in a security context should not be treated as a simple technical solution, as algorithms are infused with the values of the people who design and implement them (see Bellanova and de Goede, 2022; Hannah-Moffat, 2018). Furthermore, making judicial decisions based on predictions, rather than concrete evidence, can lead to a situation in which people are labeled as suspects based on their patterns of behavior, even if they have not actually broken any laws (see Zedner and Ashworth, 2019). This is echoed by the argument that 'low probability, high consequence' events have become increasingly central to security discourse and practices, leading to a pervasive politics of possibility, driven by a desire to regulate and control based on mere possibilities (see Amoore, 2013).

In summary, contemporary prediction has been shaped by digitalization and has permeated various fields where anticipation and prevention have become obligatory passage points. In addition to the various affordances of digital tools (see Affordances by Wood and Arpke-Wales), there have been numerous concerns about an indiscriminate application of a predictive logic. Indeed, crime has been an area where this is most pronounced.

Predictions in the criminological realm

Predictions and predictive instruments have always been important to decision-making in the context of criminal justice. As methods of quantification evolved, so did the use of sophisticated statistics in criminal justice to improve crime prediction (Berk, 2012). In the penal realm, the pursuit of statistically valid and predictively useful risk factors for recidivism and parole violations became important.

In the 1970s, researchers began to develop predictions for a range of criteria (Farrington and Tarling, 1985). Predictive methods were applied to make criminal justice more manageable through selective incapacitation of dangerous individuals. Prediction studies also dealt with predicting future rates of arrest, imprisonment, crime, and criminals. In the 1980s, the trend was towards 'actuarial justice,' where statistical techniques from insurance and risk management became part of the penal system for assessing the risk of offenses and recidivism (Feeley and Simon, 1992). Algorithms are widely used in criminal justice systems (Kehl et al., 2017). Pre-trial risk assessment instruments assess the likelihood that a defendant is a threat to public safety or will not appear in court (Brayne and Christin, 2021). During sentencing, they can be used to determine sentencing, and after they are used to predict recidivism (see Sentencing and Risk Assessments by Ugwudike). The risk scores are also used in the correctional system to determine the security classification of inmates (Mehozay and Fisher, 2018). These interventions are highly intrusive and require a high degree of accuracy.

An early example of general prediction of crime trends is Georgette Bennett's book *Crimewarps: The Future of Crime in America* from the 1980s. Few of the trends predict-

ed for the next 20 to 50 years came true. More recently, Police Chief William J. Bratton and the Los Angeles Police Department have been credited with developing the predictive policing model. In 2008 Bratton spoke widely in different fora about the successes of the Los Angeles Police Department due to the department's introduction of predictive analytics to anticipate gang violence and to support real-time crime monitoring (Brayne, 2021). In subsequent years, we encounter a plethora of terms in academic literature that describe and promote law enforcement in the age of big data (see Big Data by Završnik), such as data driven policing, big data policing, intelligence-led policing, and digital policing. These refer to the process whereby law enforcement in many Western countries has started to implement new forms of organization and knowledge production, and increasingly adopted the language and methods of computing for the purposes of prediction.

The attendant transformations are significant, and there are numerous operational contexts in which prediction and forecasting can play a role in relation to criminal justice and law enforcement.

(I) Criminal justice and law enforcement has become more dependent on digital systems for the purposes of decision-making, resource allocation, increasing operational efficiency, and security management.
(II) Predictions have become pertinent for scholars making recommendations for use in criminal justice, penal policy, and police.
(III) Predictions play a role in shaping knowledge and practice in the criminal justice system. For example, they influence the intelligence that police officers work with (Shapiro, 2019), altering crime solving practices, and introducing pre-emptive and intelligence-led tactics that rely on predictions based on crime data.

While all the above transformations are significant, the final point requires further attention. Empirical studies have made scholars sensitive to the intricacies of how knowledge is generated in the context of predictive policing (see Brayne, 2021; Egbert and Leese, 2021). First, when thinking about prediction of crime and criminal activity, one must contend with the fact that predictions can be used to target both places (hot spots) and individuals (profiling) (Hälterlein, 2021). In the case of places and phenomena, the predictions are built upon assumptions about the specific environment and the vulnerabilities that it has. This may be based on both historical and sociodemographic data. In the case of people, the predictions are based on the identification of behavioral patterns that are characteristic of a particular group of people.

Second, the way data are turned into actionable intelligence requires careful attention as various parallel processes affect how data is generated, organized, and utilized in the context of predictions (see Kaufmann et al., 2020). The contexts and formats in which data are produced and the purposes for which databases are built shapes what kind of information is available and the kind of patterns an analyst can discern. This is further determined by the way data is entered, processed, and standardized, which frequently involves human labor and judgment. Thus, the way that data have been generated and employed can and should be interrogated. The risk is that data and the pre-

dictive tools can reify biases, undermine individual freedoms, and lead to excessive criminalization of certain geographical areas and groups.

Third, prediction can be based on different epistemological approaches. In the case of theory-driven predictions, theoretical explanations, and assumptions about places ('hotspots') and the behavior of human individuals in specific circumstances (routine activity theory) provide accounts of how a particular interaction of social, physical, and behavioral factors leads to criminal activity. Theory-driven predictions tend to be more intelligible to a criminologist, and rely on subject-specific criminological knowledge, providing an account that is explainable in criminological terms. Approaches can move away from a reliance on traditions attempting to explain human behavior (social science) and borrow from alternative traditions, such as epidemiology and earthquake theory (see Hälterlein, 2021).

Conversely, criminological theory and knowledge are believed to be less relevant in the context of predictions that are based on machine learning and are generated without direct input from theory (cf. Chan and Bennett Moses, 2016). Predictions are allegedly provided by models iteratively building upon historical data to provide the most accurate prediction, while being conceptually indifferent to what they are modeling and predicting (Hälterlein, 2021). This, however, may overlook instances where algorithms are modified by situationally adding parameters that derive from causal of theoretical explanations proposed by an analyst (see Kaufmann, 2019). Thus, while digital tools and algorithms have a more pronounced role in the identification of patterns and the generation of predictions, human agency and expert judgment also play a role. However, the influence of human analysts tends to be downplayed to separate the algorithmic approach from the theory-driven approach and highlight the role of the digital component.

Defining and problematizing prediction, therefore, requires that we study moments, practices, and technologies that bring data into being and use them to generate predictions that are subsequently acted upon in correctional service or police work. This leads to questions about agency (human or algorithmic; see Agency by Krasmann), expertise, and broader considerations about the foundations of prediction in the context of the criminal justice system. Predicting deviations from the norm is paradoxical. Several risk assessment systems aim to predict dubious phenomena such as violent extremism and screen out people with serious mental health problems. It is also important to recognize that prediction in this context is a heterogeneous phenomenon. For example, different technologies and tools have been associated with enhancing the predictive and forecasting capacities of the police. What is more, predictive policing has led police authorities to become more aware of the value of the vast data at their disposal, and this has in turn led to growing interest in data integration and analysis platforms (see Egbert, 2019). Consequently, one could argue that prediction is currently distributed in a wide set of practices that do not necessarily correspond to our traditional imaginary of police work, and this has given prediction a diffused and contested character in the context of policing.

Common criticisms of predictive systems are related to accountability and a lack of transparency regarding the sociomaterial practices underpinning prediction (see Meijer and Wessels, 2019). In short, the issue is the general opacity and obscurity of the way results and outputs of predictive tools are generated and interpreted (by relying upon preconceptions, biases, and reifying existing forms of discrimination). Furthermore, the identification of social values embedded in predictive technologies focusing on crime suggests the possibility that structural bias may be implicit in other areas as well (e.g., healthcare systems (Obermeyer et al., 2019); the justice system (Ugwudike, 2020)). Opening the black box of concepts such as prediction allows us to follow what happens to police organizations and correctional systems on a practical level and what kind of societal impact this has.

Consequently, to understand prediction as an epistemic object and in practice, we consider the notion of boundary objects to be useful. We contend that prediction is malleable enough to be adopted and adapted by several criminal justice actors, yet robust enough to maintain a common identity and concept across different organizational sites and applications (cf. Star and Griesemer, 1989). This line of thought provides us with an analytical framework for simultaneously understanding predictive (but also digital, big data, data-driven) technologies as complex, changing, and stabilized entities. At the same time, it draws our attention to how prediction is generated, enacted, and managed by various stakeholders. Treating prediction as a boundary object allows us to acknowledge that prediction is part of multiple sociotechnical worlds. We can explore how prediction facilitates epistemological communication between the construction of crime and the practice of (pre-emptive) policing. Likewise, it allows the analysis to explore the different identities that prediction acquires in each social world that it inhabits: a quantum leap in the future, a threat to privacy and democracy, a significant change in police practices, a dystopian futuristic development, and so on. In this way, we can attend to the different mutations of prediction as concrete and abstract, simultaneously good and evil, futuristic and dystopian.

Conclusion

Prediction has a long history in the social sciences, and digitalization has been conducive to a renewed sense of optimism about the ability to produce meaningful predictions. Criminology, too, is part of that history by studying models for predicting criminality, offending, recidivism, as well as discussing these critically.

- The police, correctional service and criminal courts have become more reliant on digital systems of prediction in a wide variety of contexts and applications, which shapes these institutions and the knowledge that they work with.
- While digitalization has increased the ability of practitioners in various fields to analyze data and generate predictions, an indiscriminate application of predictive and anticipatory approaches in the context of the penal realm and law enforcement should be treated with caution.

- Critics have raised concerns as to how data is generated, organized, and used as input for predictive tools and technologies, and how the results are interpreted in the context of criminal justice
- There are different epistemological approaches to prediction.
  - Theory-driven predictions tend to rely on subject-specific criminological knowledge.
  - Predictions generated by object-agnostic models or algorithms are depicted as conceptually indifferent to what they are modelling and predicting, though they can rely on theoretical input.
- Treating prediction as a boundary object or product of boundary work enables us to simultaneously explore the multiple sociotechnical worlds that prediction is part of, as well as the different identities that prediction acquires in each social world that it inhabits.

The notion of prediction highlights how digital tools are transforming the nature of knowledge and expertise within the criminal justice system. Exploration of the creation and management of a boundary object such as predictive policing and risk assessments in sentencing or correctional service is of great value. It allows us to critically follow and understand the process of achieving and maintaining coherence across intersecting social contexts as well as organizations such as the police and criminal court and draws our attention to the actors implicated in this process.

Suggested reading

Brayne, S. (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Hannah-Moffat, K. (2018). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, 23 (4), 453 – 470.

Hälterlein, J. (2021). Epistemologies of predictive policing: Mathematical social science, social physics and machine learning. *Big Data & Society*, 8(1).

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.

References

Amoore, L. (2013). *The Politics of Possibility: Risk and Security Beyond Probability.* Durham, NC: Duke University Press.

Aradau, C., & Blanke, T. (2017). Politics of prediction: Security and the time/space of governmentality in the age of big data. *European Journal of Social Theory*, 20(3), 373 – 391.

Bellanova, R., & de Goede, M. (2022). The algorithmic regulation of security: An infrastructural perspective. *Regulation and Governance*, 16, 102 – 118.

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*, 28(7), 806 – 822.

Berk, R. (2012). *Criminal Justice Forecasts of Risk: A Machine Learning Approach.* New York: Springer-Verlag.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977 – 1008.

Brayne, S. (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 68(3), 608 – 624.

Brynjolfsson, E., Jin, W., & McElheran, K. (2021). The power of prediction: Predictive analytics, workplace complements, and business performance. *Business Economics*, 56, 217 – 239.

Cevolini, A., & Esposito, E., (2020). From pool to profile: Social consequences of algorithmic prediction in insurance. *Big Data & Society*, 7(2).

Chan, J., & Bennett Moses, L. (2016). Is Big Data challenging criminology? *Theoretical Criminology*, 20(1), 21 – 39.

Egbert, S. (2019). Predictive policing and the platformization of police work. *Surveillance & Society*, 17(1/2), 83 – 88.

Egbert, S. and Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London: Routledge.

Farrington, D., & Tarling, R. (eds.) (1985). *Prediction in Criminology.* New York: State University of New York.

Feeley, M., & Simon, J. (1992). The new penology; notes on the emerging strategies for corrections. *Criminology*, 30(4), 449 – 475.

Foucault, M. (2003). *Society Must Be Defended: Lectures at the Collège de France, 1975 – 1976.* New York: Picador.

Halpern, O. (2015). *Beautiful Data: A History of Vision and Reason since 1945.* Durham, NC: Duke University Press.

Hälterlein, J. (2021). Epistemologies of predictive policing: Mathematical social science, social physics and machine learning. *Big Data & Society*, 8(1).

Hannah-Moffat, K. (2018). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, 23(4), 453 – 470.

Hardyns, W., & Rummens, A. (2017). Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European Journal on Criminal Policy and Research*, 24(3), 201 – 218.

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.

Kaufman, M. (2019). Who connects the dots? Agency and agents in predictive policing. In M. Hoijtink & M. Leese (eds.), *Technology and Agency in International Relations* (pp. 141 – 163). London: Routledge.

Kaufmann, M., Thylstrup, N. B., Burgess, J. P., & Rudinow Sætnan, A. (2020). Data criticality. *STS Encounters – DASTS working paper series*, 11(1), 227 – 254.

Kehl, D., Guo, P., & Kessler, S. (2017). *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing.* Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

Kersing, M., van Zoonen, L., Putters, K., & Oldenhof, L. (2022). The changing roles of frontline bureaucrats in the digital welfare state: The case of a data dashboard in Rotterdam's Work and Income department. *Data & Policy*, 4(e24).

McCulloch, J., & Wilson, D. (2016). *Pre-crime: Pre-emption, Precaution and the Future.* London: Routledge.

Mehozay, Y., & Fisher, E. (2018). The epistemology of algorithmic risk assessment and the path towards a non-penology penology. *Punishment & Society*, 21(5), 523 – 541.

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, 42(12), 1031 – 1039.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447 – 453.

Oster, J., (2021). Code is code and law is law—the law of digitalization and the digitalization of law. *International Journal of Law and Information Technology*, 29(2), 101 – 117.

Porter, T. M., & Ross, D. (eds.) (2008). *The Cambridge History of Science, Volume 7: The Modern Social Sciences.* Cambridge: Cambridge University Press.

Rehman, A., Naz, S., & Razzak, I. (2022). Leveraging big data analytics in healthcare enhancement: Trends, challenges and opportunities. *Multimedia Systems*, 28(4), 1339 – 1371.

Scheel, S., Ruppert, E., & Ustek-Spilda, F. (2019). Enacting migration through data practices. *Environment and Planning D: Society and Space*, 37(4), 579 – 588.

Shapiro, A. (2019). Predictive policing for reform? Indeterminacy and intervention in big data policing. *Surveillance & Society*, 17(3/4), 456 – 472.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907 – 39. *Social Studies of Science*, 19(3), 387 – 420.

Ugwudike, P. (2020). Digital prediction technologies in the justice system: The implications of a 'race-neutral' agenda. *Theoretical Criminology*, 24(3), 482 – 501.

Zedner, L. (2009). *Security.* London: Routledge.

Zedner, L., & Ashworth, A. (2019). The rise and restraint of the preventive state. *Annual Review of Criminology*, 2, 429 – 450.

Lee A. Bygrave

43 Privacy and data protection

Abstract: This entry elucidates the semantics of privacy and data protection, taking account of transnational variations in how they are understood. Their interrelationship is also explained as is their relationship to closely linked concepts, such as data security and bio-privacy. Finally, the entry briefly canvasses legal and social norms for safeguarding privacy and personal data.

Keywords: privacy, data protection, data security, secrecy, surveillance, law

Introduction

Privacy is a term about which there has been a long and ultimately unresolved debate concerning its proper meaning. Literature devoted to its semantics is massive (for useful overviews, see DeCew, 1997; Solove, 2008; Koops et al., 2017). Much of the literature is philosophical and legal but contains significant input from other disciplines. These include sociology, psychology, and criminology, along with their more specialized offshoots, such as surveillance studies. The debate over the semantics of privacy escalated during the 1980s and 1990s, largely in conjunction with discussion over the societal effects of increasing mass surveillance (see Surveillance by Lyon)—i. e., the systematic monitoring of ever greater numbers of persons and ever larger amounts of personal data with a view to exercising control or generating profit (see, e. g., Flaherty, 1989; Lyon and Zuriek, 1996; Lyon, 2007; Zuboff, 2019). Privacy has been flagged as a key interest (or congeries of interests) that is seriously threatened by such development, particularly in the context of crime control and surveillance capitalism. Yet, somewhat paradoxically, arriving at a precise, universally agreed definition of privacy has proven well-nigh impossible. This is partly because of the nebulous, loose, and somewhat haphazard way in which the term is often used, and partly because it is multifaceted—as elaborated below.

The semantics of data protection have also come to the fore in conjunction with discussion concerning mass surveillance. Like privacy, data protection is typically pitched as an important bulwark for upholding civil liberties in the face of state and business interests in exploiting personal data for various ends—national security, crime control, organizational efficiency, profit generation, etc. However, the semantics of data protection are far easier to navigate than those pertaining to privacy. Broad consensus exists as to the core meaning of data protection, and the legal-regulatory context in which it is applied is significantly more bounded than is the case for privacy. Whereas privacy denotes an interest (or set of interests) that potentially range(s) across all existential contexts, data protection denotes a set of norms and other regulatory measures pertaining to the processing of a particular type of data. Nonetheless, use of the term presents challenges. For persons who are unfamiliar with it, data protec-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-044

tion on its face may be easily conflated with data security and similar security-specific terms such as computer security and cybersecurity, yet, as shown below, data protection is not fully commensurate with security of data. Further, the term is anemic in the sense that it fails to flag the interests and values served by the norms and measures it denotes. This has led to considerable debate over the rationale of data protection as a regulatory endeavor (see, e.g., Bygrave, 2002; Bygrave, 2025). Part of that debate is the nature of the connection between privacy and data protection—in particular, the degree to which data protection is predominantly concerned with safeguarding privacy.

Privacy semantics in greater depth

Defining privacy is a challenging and contentious endeavor. Adding to the challenge is that many non-English languages do not operate with direct counterparts of 'privacy.' For instance, a commonly used equivalent for privacy in Chinese is 'yinsi,' which connotes a shameful secret (Lü, 2005). However, privacy is not the same as secret, which may be defined in terms of "intentional concealment" (Bok, 1982: 5), nor is privacy necessarily shameful. To take another example, the closest equivalents to privacy in Norwegian are 'privatlivet' ('private life') and 'privatlivets fred' (the peace of private life): neither term is fully commensurate with all the various possible connotations of privacy, as shown below. Thus, persons who do not have English as their native language may struggle to get a complete sense of privacy's many facets.

A related challenge springs from the ideological underpinnings of privacy-focused discourse, which is most extensively developed in Western cultures with strong liberalist traditions. As Lukes observes, privacy in the sense of a "sphere of thought and action that should be free from 'public' interference" constitutes "perhaps the central idea of liberalism" (Lukes, 1973: 62). The fact that much of the discourse on privacy as a threatened interest tends to assume the virtue of Western liberalist assumptions undercuts its global appeal. Nonetheless, there are also 'privacy skeptics' within the Western sphere who question the societal desirability of robust privacy protections —e.g., from economic, communitarian, or feminist viewpoints (see, e.g., Posner, 1978; Etzioni, 1999; Allen, 1988).

A recent comprehensive mapping of privacy's dimensions across various fields, including law, parses privacy into eight basic types (bodily, intellectual, spatial, decisional, communicational, associational, proprietary, and behavioral), with a ninth type (informational) as an overlapping but distinct overlay to the former types (Koops et al., 2017). The study is not intended to be exhaustive but underscores the huge semantic and contextual space that privacy occupies.

Sometimes, particular aspects of this space are specifically flagged by adding a prefix to 'privacy.' An example is 'bio-privacy,' which is used to denote the "privacy issues of biometric technology" (Liu, 2012: 21), thus emphasizing the bodily, behavioral, and informational dimensions. Bio-privacy concerns become relevant primarily in respect of the deployment of biometrics-based systems of identification or verification/authen-

tication—for instance, iris-scanning devices for controlling access to closed-off spaces as part of border/immigration control, or automated facial recognition cameras for identifying criminal suspects in crowds (see Biometric Failure by Magnet and Din). Another, closely related example is 'genetic privacy,' which is typically used to denote protection from non-voluntary disclosure of genetic information (see, e. g., Allen, 1997; Lunshof et al., 2008).

At the same time, it is important to note how new technologies and organizational practices are blurring or merging the various dimensions of privacy. A salient instance is the gradual ontological evisceration of the traditional distinction between the physical/physiological human body as such and information about the body, brought on by the increased focus on the body as a source of information, particularly due to developments in genetic testing and biometrics. In this respect, van der Ploeg refers to an emergent "ontology of informatized bodies" (van der Ploeg, 2005). There is growing legal recognition of this trend, exemplified in the willingness of courts and legislators to treat human biological material, such as DNA (deoxyribonucleic acid), as personal data subject to data protection norms (see DNA/Big Genome Data by Kaufmann). The 2008 judgment of the European Court of Human Rights in *S and Marper v. United Kingdom* is a case in point, with the Court stating that cellular samples from persons who were apprehended by the police constitute personal data for the purposes of the Council of Europe Convention on Data Protection (see further Bygrave, 2010: 7–8).

Scholarly conceptions of privacy frequently have differing focal points in respect of the term's radius. One conception pitches privacy predominantly in terms of non-interference. The most famous example in point is Warren and Brandeis' definition of the right to privacy in US law as the "right to be let alone" (Warren and Brandeis, 1890). Another conception pitches privacy as a state of limited accessibility (e. g., Gavison, 1980; Bygrave, 2002). A third conception of privacy is as a claim to information control. Westin's definition of privacy is a landmark instance of this view: "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1970: 7). A fourth conception connects privacy exclusively to intimate or sensitive facets of persons' lives. Upon this view, disclosure of non-intimate or non-sensitive information about a person does not diminish their privacy (e. g., Inness, 1992; Wacks, 1989). Yet another conception brings context to the fore. A key instance is Nissenbaum's framing of privacy in terms of 'contextual integrity,' whereby privacy is essentially concerned with maintaining respect for the structure of social settings (Nissenbaum, 2010). In this perspective, privacy speaks to the preservation of formal and informal social norms as to what behaviour (e. g., regarding information disclosure) is appropriate in particular spheres of life or activity. For instance, privacy would uphold the reasonable expectation that the norms for information disclosure in bedroom activity differ significantly from the norms for information disclosure in publicly accessible places.

The aforementioned list of conceptions is not exhaustive but is sufficient to convey the variation of perspectives on privacy's semantic core. From the list, it becomes evident that there are also differences over whether privacy in itself is best understood as

a state of being or as a normative position, such as a moral claim or legal right, and over whether privacy is a form of autonomy or a completely independent category of state or interest.

The semantics of data protection in greater depth

The term 'data protection' derives from the German term 'Datenschutz.' The latter has roots in terminology denoting security of data (namely, 'Datensicherung' and 'Datensicherheit'). However, data protection and its German equivalent were coined in order to flag a set of measures and concerns that go well beyond simply ensuring data security (Simitis, 2006: 62–63).

Data protection denotes a set of measures specifically aimed at regulating the processing of data relating to, and facilitating identification of, persons (i.e., personal data) in order to safeguard, at least partly, the privacy and related interests of those persons. The measures embody a set of largely procedural principles addressing, *inter alia*, the purposes of the data processing and the quality of the data, and stipulating measures to ensure that the processing is transparent to, and capable of being influenced by, the person to whom the data relate ('data subject'). In most countries, these principles and measures have been set into legislation, and the resultant corpus of law tends accordingly to be termed 'data protection law.' The EU's General Data Protection Regulation (GDPR) of 2016 is a landmark example in point.

However, the nomenclature for this sort of law is not uniformly termed 'data protection.' In certain countries outside Europe, such as the US, Australia, Canada, and New Zealand, this body of law tends to be labeled 'privacy law,' 'information privacy law' or, more recently, 'data privacy law.' This labeling reflects in part the central place of 'privacy' in these countries' public debates over the societal implications of computerized processing of personal data (see Data Justice by Redden). For instance, when these debates first took off in the US during the 1960s, privacy was invoked as a key term for the interests that were perceived to be threatened (Westin, 1970). At the same time, the semantics of privacy were (and remain) sufficiently broad and flexible to address what was then (and still is) regarded as a basic danger of computer (mis)use, namely the increased potential for large organizations to amass data on individuals and thereby subject them to excessive control (Bygrave, 2002, 2025). Even in Europe, the right to respect for private life is frequently construed and applied as a weapon of data protection, also by the EU Court of Justice and the European Court of Human Rights (Bygrave, 2025). Further, the notion of 'informational self-determination' has become a leitmotif for much of data protection law, and this notion fits comfortably with Westin's conception of privacy as a form of information control (Westin, 1970). Thus, there are close ties between privacy and data protection.

Nonetheless, it bears emphasis that data protection and privacy are not fully commensurate with each other. Data protection law embraces more than what are typically regarded as privacy concerns (Bygrave, 2025). And, as elaborated above, privacy as such

has various dimensions (spatial, bodily, etc.) with which data protection tends not to deal directly.

Similarly, data protection is not fully commensurate with data security. Although data security is a component of data protection, the latter embraces other rules and measures too. Moreover, data security on its own may serve a broader range of concerns than data protection: whereas a central aim of data protection is safeguarding data subjects' privacy-related interests, data security as such can also be aimed at safeguarding the interests of users of all kinds of data (not just personal data) in the name of, say, national security. The same applies with the overlapping areas of information security, information systems security, and cybersecurity. The security measures are mainly directed towards ensuring that data are processed in line with the expectations of those who steer or use a given information system. The chief sub-goals for these measures are maintenance of the confidentiality, integrity, and availability of information in an information system as well as appropriate protection of the system itself, wherein 'confidentiality' indicates that the information and system are protected from unauthorized disclosure, 'integrity' describes their protection from unauthorized modification, and 'availability' denotes that they are accessible and usable on demand by authorized actors, systems, or programs (Bygrave, 2021). In many cases, these measures may serve to promote data protection, but they can obviously come into conflict with the latter as well.

Implementation of privacy and data protection measures

Protection of privacy and personal data is achieved through a variety of means. Public attention tends to be directed towards high-profile legislative measures, such as the GDPR, which constitutes, in the eyes of many, the most globally influential 'gold standard' of its kind (Bradford, 2020). This perception is partly due to the GDPR's relatively powerful sanctions and enforcement regime. However, the strength of that regime has proven to be significantly greater 'on paper' than in practice (see, e. g., Gentile and Lynskey, 2022), thus underlining (yet again) that the classic gap between 'law in books' and 'law in practice' tends to be extremely difficult to bridge. Concomitantly, the level of privacy and data protection 'on the ground' tends only to be partly a function of legislative or judicial efforts; other measures often play an equally or more significant role. These include soft law instruments in the form of technical standards, recommendations, codes of practice, and the like. Yet, arguably, the most enduring constituents of the societal bedrock of privacy and data protection are more informal or intangible behavioural norms rooted in context-sensitive notions of social propriety.

Conclusions

- Although signifying important determinants of the quality of societal life, both privacy and data protection remain contested, nebulous concepts.
- The semantics of privacy are especially difficult to comprehend due to the disparate ways in which the term tends to be invoked and the multiplicity of its dimensions.
- Nonetheless, claims that privacy and data protection are too vacuous to be analytically or normatively useful should be rejected; although they both suffer from terminological imprecision, they each have a sufficiently tangible core of meaning to warrant their continued use in legal-regulatory policy and other public discourse.

Suggested reading

Bygrave, L. A. (2025). *Data Privacy Law: An International Perspective.* 2nd ed. Oxford: Oxford University Press.

Koops, B.-J., Newell, B. C., Timan, T., Škorvánek, I., Chokrevski, T., & Galič, M. (2017). A typology of privacy. *University of Pennsylvania Journal of International Law*, 38, 483 – 575.

Solove, D. (2008). *Understanding Privacy.* Cambridge, MA: Harvard University Press.

References

Allen, A. (1988). *Uneasy Access: Privacy for Women in a Free Society.* Totowa, NJ: Rowman & Littlefield.

Allen, A. (1997). Genetic privacy: Emerging concepts and values. In M. A. Rothstein (ed.), *Protecting Privacy and Confidentiality in the Genetic Era* (pp. 31 – 59). New Haven: Yale University Press.

Bok, S. (1982). *Secrets: On the Ethics of Concealment and Revelation.* New York: Pantheon.

Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World.* Oxford: Oxford University Press.

Bygrave, L. A. (2002). *Data Protection Law: Approaching its Rationale, Logic and Limits.* The Hague: Kluwer Law International.

Bygrave, L. A. (2010). The body as data? Biobank regulation via the 'back door' of data protection law. *Law, Innovation and Technology*, 2, 1 – 25.

Bygrave, L. A. (2021). Security by design: Aspirations and realities in a regulatory context. *Oslo Law Review*, 8(3), 126 – 177.

Bygrave, L. A. (2025). *Data Privacy Law: An International Perspective.* 2nd ed. Oxford: Oxford University Press.

DeCew, J. W. (1997). *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology.* Ithaca, NY: Cornell University Press.

Etzioni, A. (1999). *The Limits of Privacy.* New York: Basic Books.

Flaherty, D. H. (1989). *Protecting Privacy in Surveillance Societies.* Chapel Hill, NC: University of North Carolina Press.

Gavison, R. (1980). Privacy and the limits of law. *Yale Law Journal*, 89, 421 – 471.

Gentile, G., & Lynskey, O. (2022). Deficient by design? The transnational enforcement of the GDPR. *International & Comparative Law Quarterly*, 71(4), 799 – 830.

Inness, J. C. (1992). *Privacy, Intimacy, and Isolation.* Oxford: Oxford University Press.
Koops, B.-J., Newell, B. C., Timan, T., Škorvánek, I., Chokrevski, T., & Galič, M. (2017). A typology of privacy. *University of Pennsylvania Journal of International Law*, 38, 483 – 575.
Liu, N. Y. (2012). *Bio-Privacy: Privacy Regulations and the Challenges of Biometrics.* London: Routledge.
Lü, Y.-H. (2005). Privacy and data privacy issues in contemporary China. *Ethics and Information Technology*, 7, 7 – 15.
Lukes, S. (1973). *Individualism.* Oxford: Blackwell.
Lunshof, J., Chadwick, R., Vorhaus, D., & Church, G. M. (2008). From genetic privacy to open consent. *Nature Reviews Genetics*, 9, 406 – 411.
Lyon, D. (2007). *Surveillance Studies: An Overview.* Cambridge: Polity Press.
Lyon, D., & Zuriek, E. (1996). *Computers, Surveillance, and Privacy.* Minneapolis: University of Minnesota Press.
Nissenbaum, H. (2010). Privacy in Context: Technology, Policy, and the Integrity of Social Life. Redwood City, CA: Stanford University Press.
Posner, R. A. (1978). The right to privacy. *Georgia Law Review*, 12, 393 – 422.
Simitis, S. (ed.) (2006). *Bundesdatenschutzgesetz.* 6th ed. Baden-Baden: Nomos Verlagsgesellschaft.
Solove, D. (2008). *Understanding Privacy.* Cambridge, MA: Harvard University Press.
van der Ploeg, I. (2005). *The Machine-Readable Body: Essays on Biometrics and the Informatization of the Body.* Maastricht: Shaker Press.
Wacks, R. (1989.) *Personal Information: Privacy and the Law.* Oxford: Clarendon Press.
Warren, S., & Brandeis, L. (1890). The right to privacy. *Harvard Law Review*, 4, 193 – 220.
Westin, A. F. (1970). *Privacy and Freedom.* New York: Atheneum.
Zuboff, S. (2019). *The Age of Surveillance Capitalism.* London: Profile Books.

Heidi Mork Lomell

44 Privatization

Abstract: Almost every aspect of the digital is private. To understand digitization, we therefore need to understand privatization. This chapter outlines some of the aspects of privatization by exploring major differences between a public and a private approach to criminalization, crime, and control.

Keywords: privatization, private space, regulation, commercial logics

Introduction

On January 6, 2021, a mob of Donald Trump's supporters attacked Capitol Hill in an attempt to prevent the US Congress from approving the victory of president-elect Joe Biden. After the attacks, Facebook, Instagram, Twitter, Snapchat, and YouTube suspended Donald Trump's accounts because of the role his social media posts played in inciting the violence and also due to the risk of inciting further attacks (Mamak, 2023). This was not a result of a police investigation, a criminal charge, or a court verdict. Suspension from social media is not a punishment for a crime committed, but a result of a decision made by the social media companies for breaking their terms of services. To some, the suspensions demonstrated that no one is above the law. For others, the case illustrates that we all are, just as Donald Trump was, subject to the arbitrary and unaccountable power of companies (Floridi, 2021).

Much of our personal, social, and professional life has migrated onto private platforms (see Platforms by Egbert). Both private citizens and public institutions are increasingly dependent on a privatized, digital infrastructure (Andrejevic, 2013). Almost every aspect of the digital is private: private companies invent, design, produce, sell, maintain, and in many respects even *govern* the digital (Floridi, 2020).

This raises several questions for criminologists. Criminalization, crime, and crime control have traditionally been understood and studied as public issues. These are fields inhabited by public and political actors such as lawmakers, bureaucrats, police officers, judges, and prison guards. What happens when private and commercial interests enter these fields through digitization? How do we approach the role of private interests in digital criminology?

The private foundation of digital technologies is a key component for studying crime and crime control in digitized societies. To understand digitization, we need to

Disclaimer and acknowledgments: The chapter is written as part of the CRIMKNOW project (Private knowledge, public issues: Digitalization and private economies of knowledge in criminal justice), financed by the Research Council of Norway (project no. 334953).

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-045

understand privatization, especially some of the major differences between a public and a private approach to our core topics: criminalization, crime, and control.

In this entry, I will explore how the concept of privatization offers an important point of departure for understanding digital criminology. I will start by taking a closer look at the concept itself and its counterpart: the private and the public. I will then explore what is meant by privatization and how privatization has been studied in criminology, before addressing some of the major issues in digital criminology through the lens of privatization.

Private and public

The categories of public and private are central in structuring our social lives (d'Entrèves and Vogel, 2000: 1). 'Public' is associated with the state, with politics, and with the community, while 'private' is associated with the economy, the market, but also the home, the family, and the personal. We use these categories to distinguish between sectors, such as the state and the economy, the political and the market, and the government and the private citizen. We also use them to distinguish between realms or spaces, either literally, for instance land or property, or more figuratively, for instance information or data. A public space is in principle accessible and open for all, while private space has limited access. In some countries, information such as your annual income or your criminal record is considered public, while in other countries this information is considered private. Some data are public, such as the crime statistics, which is produced by the police and the courts, while other data are private, for example data from social media platforms on search patterns.

What is private and what is public varies in time and space. And even if we can view the concepts as binaries, there are and have always been overlaps. For example, private markets are not entirely free from public regulation, and the private home is not totally secluded from public interventions—whether it is regulations and inspections concerning electricity and fire protection, or the child protection services inspecting a home due to a report of concern.

Even before digitization, the demarcation between public and private was blurry. With digitization, such distinctions have become even more complex. Private data or information can easily be made public, either by consent or against a person's will. One example is how we increasingly both document and share our private life online, and thereby to broader publics than before.

Some private social media platforms are defined as a public space by law. Our 'private' messages or posts on social media platforms can therefore be defined and treated by law as 'public'; offensive comments made on private social media accounts have led to prison sentences (Rustin, 2014). Even on platforms that require a login to access, private information, communication, and images risk becoming public through 'sharing' functions (Powell et al., 2018: 28). In 2018, more than 1,000 young people in Denmark were charged for the possession and sharing of child pornography, indecency, and

wrongfully disclosing of private affairs after sharing two video sequences containing sexual material involving persons under the age of 18 on the Facebook chat-platform Messenger (Holst and Adolphsen, 2020). These examples illustrate that our understanding of what is defined as private and public online is still developing.

Another complicating factor is that these platforms are developed, owned, administrated, and regulated by private corporations. This means that even if we choose to be or are defined by law as 'public' in our online life, we are literally on private grounds and subject to their norms and regulations.

Privatization

Privatization means that public goods, services, and/or space are overtaken by private companies. During the 1990s a range of network utilities, from telecommunications, power supplies to railways, were privatized in many countries (Offer, 2022). In its purest form privatization implies the total transfer of public assets to the private sector as manifested in the selling off of state-owned businesses, industries, and utilities (Hamerton and Hobbs, 2022: 4). However, the concept is used to refer to a wide range of processes, from the sale or contracting out of public institutions to private companies, to various forms of deregulation (Matthews, 1989: 1).

In some cases, privatization implies transferring industries and/or services *back* to the private sector: The growth of the social welfare state implied the public taking over the responsibility of the welfare of the people. In other words, what preceded the privatization of the public sector was a 'publicization' of private services with the rise of, for example, public social security, public health services, public transport, public schools, etc. Even policing functions were 'private' before the establishment of the modern, public police (Zedner, 2006). Crime was a branch of torts, where the victim had to bear the costs of law enforcement and prosecution, and compensation dominated over retribution (Feeley, 2018).

Privatization in criminology and criminal justice

While privatization was the buzzword of the 1980s and 1990s, proposals to privatize parts of the criminal justice system were met with "a widespread sense of bemused disbelief" as it was widely assumed that "the provision of laws, punishment and crime control constitute a unique and privileged realm of activity that should be provided by the State" (Matthews, 1989: 1). The monopoly over the use of legitimate force is one of the essential characteristics of a state. Nevertheless, Matthews and colleagues documented that there already existed a significant degree of privatization within the criminal justice system.

Since the 1980s, criminologists gained considerable insight into the emergence and impact of new configurations of the state and the private sector in certain areas of

crime and security governance—often framed as the rise of neo-liberal forms of governance (Crawford, 2006). In his seminal work *Crime Control as Industry* in 1993, Christie documented how the growth in prison sentences in the US was closely related to the privatization of prisons (Christie, 2000). Within policing and security studies, Clifford Shearing and Philip Stenning published several studies in the 1980s and set the ground for a flourishing research tradition on both private security providers and the privatization of public policing (Johnston, 1992; Jones and Newburn, 1998; Shearing and Stenning, 1981, 1985, 1987; Wakefield, 2003). The increasing privatization of crime control and security continue to draw attention within criminology and criminal justice studies, however, often neglecting the digital aspects of the development (Daems and Beken, 2018; Hamerton and Hobbs, 2022; Hucklesby and Lister, 2018; Simmons, 2023).

The growth in private policing and private security has been connected to the parallel growth in 'mass private property,' which is privately owned land that is open or partly open to the public, often policed by private regulation and private security services. Examples include shopping malls, airports, amusement parks, and gated communities. The rules and regulations that govern such places are not primarily the criminal law, but rather (or in addition), corporate house rules and property law. Research has shown that these privately owned, semi-public or public places are policed not for the enforcement of criminal law, but rather for the purposes of both restricting public access and maximizing profit (Shearing and Stenning, 1981, 1985). Private security companies do not uphold public order and the rule of law, but the private order of their clients (Shearing, 1992).

Criminological research on the consequences of privatization has strong links to surveillance studies (Lyon, 2007). Both fields have studied the implications of privatization on issues on both social and crime control and more specifically on surveillance practices (see Surveillance by Lyon). The findings implicate that privatized public space is controlled by other norms, with other means and with other consequences than public space controlled by public police (Lomell, 2004). Private control is based on an instrumental, not moral, foundation, where order is whatever maximizes profit. The instrumental language of profit and loss replaces the moral language of criminal justice. Control becomes consensual, and rule-breaking leads to exclusion and banishment (Shearing and Stenning, 1985).

Corporate and state surveillance merge and promise both increased security from the state and increased convenience from the corporations (Giroux in Powell et al., 2018: 55). These differences between corporate and state surveillance resembles earlier studies of the logic of private control. Shearing and Stenning pointed to the instrumental and commercial underpinnings of privatized control instead of the moral basis of criminal justice (Shearing and Stenning, 1985). Also, Zygmunt Bauman has highlighted the more seductive dimensions of private surveillance and control: "everything moves from enforcement to temptation and seduction; surveillance society no longer chases its victims but finds its volunteers" (in Walker-Munro, 2023: 121).

Privatization and digital criminology

The findings from criminological research on privatization are a valuable starting point for researching privatization in digitized societies. Important questions to ask are: Who makes the rules, what are they, who are the control agents, what are their powers, and what are the consequences of rule-breaking?

For digital criminology, privatization means something different than privatizing institutions and practices that were previously run by public service. In digitized societies, many private, social, public, and professional aspects of life are interlinked with privately run social media platforms and computer programs. This is new. The private sector has always been part of society, but now it is increasingly providing the digital infrastructure for both citizens and governments (see Infrastructures by Grisot and Parmiggiani). We rely upon the private sector for the provision of digital utilities and services such as communication, search engines, programs, and storage. This deserves attention. Not only is it relevant to study the consequences of building a digital society on the premises of private companies, both literarily and symbolically. Criminologists should further ask how this trend affects law-making, law-breaking, and law enforcement. Here, the commercial logics that permeate digital technologies are key. While many of us think of digital space as public—or at least semi-public, the owners remain private. The 'digital commons' are private commons where we are first and foremost consumers and products, and only secondly citizens with rights.

Most, not all, digital infrastructures have been private and commercial from the outset. Despite this, the concept privatization is useful, because it helps us to focus on how our daily life in our digital societies and on digital platforms are more and more taking place on platforms and programs governed by corporate law and regulations. This has profound consequences for crime and crime control.

One example is the growth and impact of the companies' 'terms of services' that take on the role of criminal law and procedure for the regulation of online behavior, as we saw with the suspension of Donald Trump's social media accounts. Traditionally, criminologists use criminal law and criminal justice agencies as their point of departure for studying behavior. The process of law-making, law-breaking, and law enforcement requires a law. Online, however, concepts such as 'community standards' and 'terms of services' have gained prevalence. The process of making, breaking, and enforcing them are less transparent for users but also for researchers alike. And yet, they can have severe impact for the individual. On Facebook,[1] for example, users may report posts and comments as 'inappropriate' or 'controversial.' The result of reporting might be a 'Facebook jail sentence,' [2] which means being denied posting for a certain amount of time. Such a 'sentence' is imposed by automated or real 'Facebook

1 https://transparency.meta.com/policies/community-standards/

2 https://www.online-tech-tips.com/facebook-jail-how-you-will-be-punished-for-misbehaving-on-facebook/

censors,' and might have severe consequences for people, either professionally or personally (McGoldrick, 2013). While banishment is an old form of punishment and widely used by private security companies (Lomell, 2004), 'cyber-banishment' (Mamak, 2023), or 'withdrawal of system access authorization' (R. Jones, 2000), represents the digital version, currently enforced by the platforms' private owners.

The fact that digitized public space is private at its base, then, shapes and controls the norms and behaviors of its users. One key aspect is here that companies are first and foremost focused on profit, not justice, nor democracy: "While privately-owned social media platforms are nominally facilitating democratic engagement, they are primarily optimized for data retention, advertising exposure and other profit-making strategies" (Powell et al., 2018: 54). Zuboff famously coined the term *surveillance capitalism* to describe the current situation where online data is extracted and monetized; often without consent (Zuboff, 2019).

Privatization of knowledge production

The data that users produce online are a central component of corporations' business models (Walker-Munro, 2023). They provide companies with valuable insights into their consumers. Corporations therefore routinely collect more data than strictly necessary. One example is the company TikTok that captures the location, calendar, and contacts of its users, in addition to all other apps installed on the phone. While this information appears to serve no valid purpose for delivering TikTok's services (Perkins in Walker-Munro, 2023: 122), it is valuable, as these data can be of interest to other corporations and to national or foreign security services. Another example is Cambridge Analytica, who used a Facebook quiz to collect personal information of around 87 million users and then used that information to generate political and voter insights that was then sold on to political campaigns and parties (Brayne, 2020: 24).

Private companies collect and aggregate information from both private sources and public records to make these data available to those who can pay the fee for access. The police does not only buy data collected by private companies; they also use private sector platforms to store, share, and analyze data they themselves have collected (Brayne, 2020: 25). So-called predictive policing is based on software, systems, and algorithms developed by companies such as PredPol, HunchLab, IBM and Palantir (Brayne, 2020; see Policing by Wilson).

One important but overlooked aspect of privatization in digital criminology is the privatization of knowledge production. The knowledge base of contemporary criminal justice is, due to extensive digitization and marketization, undergoing fundamental transformations that involve a greater prominence of private actors and their logics. Private and hybrid knowledge providers, such as consultancies and NGOs, depart from established traditions for transparency as well as scientific and democratic oversight in criminology and criminal justice. The role of private actors in contemporary criminal justice policy development and implementation is a topic that deserves atten-

tion from digital criminologists since they are intrinsically linked to the development and sale of digital infrastructures.

Conclusion

Criminology's core topics are first and foremost public issues. Digital criminology, with its focus on how digitization alters criminalization, crime, and crime control, will benefit from paying attention to the private actors and logics that are fundamental to digitization: "it is ultimately the private companies who regulate these spaces, determining who has access, which content is to be shared, how widely it is viewed [...] and on what grounds it is removed. In this corporatisation of both public and private life, companies such as Facebook and Google engage in practices of 'networked authoritarianism' and 'corporate censorship'" (Powell et al., 2018: 28).

It is therefore important to turn our attention to the publicness/privateness of various actors in criminal justice, but also in the privatized governance of the digital realm. A crucial task is to study similarities and differences in modes of thinking about regulation and control between public and private actors within the field of digital criminology. If the governance of the digital realm is left to private enterprises, business strategies, self-regulation and market forces, long-held criminal justice principles such as the rule of law, due process, proportionality, and a right to a fair trial, not to mention democratic oversight, accountability and transparency, risk being disregarded in increasingly crucial areas of contemporary societies (Floridi, 2021).

Main takeaways

- Almost every aspect of the digital is private: private companies invent, design, produce, sell, maintain, and in many respects even *govern* the digital.
- To understand digitization, we need to understand privatization, especially some of the major differences between a public and a private approach to our core topics: criminalization, crime, and control.
- Important questions to ask when studying digital platforms are: What data do platform owners have access to? Who makes the rules, what are they? Who are the control agents, what are their powers and what are the consequences of rule-breaking?

Suggested reading

Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing.* New York: Oxford University Press, USA.

Hamerton, C., & Hobbs, S. (2022). *Privatising Criminal Justice: History, Neoliberal Penality and the Commodification of Crime.* London: Routledge.
Simmons, R. (2023). *Private Criminal Justice: How Private Parties Are Enforcing Criminal Law and Transforming Our Justice System.* Cambridge: Cambridge University Press.

References

Andrejevic, M. (2013). Exploitation in the data mine. In C. Fuchs, K. Boersma, A. Albrechtslund, & M. Sandoval (eds.), *Internet and Surveillance: The Challenges of Web 2.0 and Social Media* (pp. 71 – 88). New York: Routledge.
Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing.* Oxford: Oxford University Press, USA.
Christie, N. (2000). *Crime Control as Industry. Towards Gulags, Western Style.* London: Routledge.
Crawford, A. (2006). Networked governance and the post-regulatory state? *Theoretical Criminology*, 10(3), 449 – 479.
d'Entrèves, M. P., & Vogel, U. (2000). Public and private: A complex relation. In M. P. d'Entrèves & U. Vogel (eds.), *Public and Private. Legal, Political and Philosophical Perspectives* (pp. 1 – 16). London: Routledge.
Daems, T., & Beken, T. V. (2018). *Privatising Punishment in Europe?* London: Routledge.
Feeley, M. M. (2018). Privatizing criminal justice: A historical analysis of entrepeneurship and innovation. In T. Daems & T. V. Beken (eds.), *Privatising Punishment in Europe?* (pp. 26 – 52). New York: Routledge.
Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology*, 33(3), 369 – 378.
Floridi, L. (2021). Trump, Parler, and regulating the infosphere as our commons. *Philosophy & Technology*, 34(1), 1 – 5.
Hamerton, C., & Hobbs, S. (2022). *Privatising Criminal Justice: History, Neoliberal Penality and the Commodification of Crime.* London: Routledge.
Holst, N. S., & Adolphsen, C. (2020). Adolescents and sexual violations online. *Juvenile and Family Court Journal*, 71(3), 43 – 55.
Hucklesby, A., & Lister, S. (eds.) (2018). *The Private Sector and Criminal Justice.* London: Palgrave Macmillan.
Johnston, L. (1992). *The Rebirth of Private Policing.* London: Routledge.
Jones, R. (2000). Digital rule. Punishment, control and technology. *Punishment & Society*, 2(1), 5 – 22.
Jones, T., & Newburn, T. (1998). *Private Security and Public Policing.* Oxford: Clarendon Press.
Lomell, H. M. (2004). Targeting the unwanted. Video surveillance and categorical exclusion in Oslo, Norway. *Surveillance & Society*, 2(2/3), 346 – 360.
Lyon, D. (2007). *Surveillance Studies. An overview.* Cambridge: Polity Press.
Mamak, K. (2023). Cyber banishment: An old sanction for virtual spaces. *Criminal Justice Studies*, 36(2), 133 – 145.
Matthews, R. (ed.) (1989). *Privatizing Criminal Justice.* London: Sage Publications.
McGoldrick, D. (2013). The limits of freedom of expression on Facebook and social networking sites: A UK perspective. *Human Rights Law Review*, 13(1), 125 – 151.
Offer, A. (2022). *Understanding the Private–Public Divide: Markets, Governments, and Time Horizons.* Cambridge: Cambridge University Press.
Powell, A., Stratton, G., & Cameron, R. (2018). *Digital Criminology: Crime and Justice in Digital Society.* New York: Taylor and Francis.

Rustin, S. (2014, June 13). Is it right to jail someone for being offensive on Facebook or Twitter? *The Guardian.* https://www.theguardian.com/law/2014/jun/13/jail-someone-for-being-offensive-twitter-facebook

Shearing, C. D. (1992). The relation between public and private policing. In M. Tonry & N. Morris (eds.), *Modern Policing* (Vol. 15) (pp. 399 – 434). Chicago: The University of Chicago Press.

Shearing, C. D., & Stenning, P. C. (1981). Modern private security: Its growth and implications. In M. Tonry & N. Morris (eds.), *Crime and Justice: An Annual Review of Research* (Vol. 3) (pp. 193 – 245). Chicago: Chicago University Press.

Shearing, C. D., & Stenning, P. C. (1985). From the panopticon to Disney World: The development of discipline. In A. N. Doob & E. L. Greenspan (eds.), *Perspectives in Criminal Law* (pp. 335 – 349). Toronto: Canada Law Book Inc.

Shearing, C. D., & Stenning, P. C. (1987). *Private Policing* (Vol. 23). London: SAGE Publications.

Simmons, R. (2023). *Private Criminal Justice: How Private Parties Are Enforcing Criminal Law and Transforming Our Justice System.* Cambridge: Cambridge University Press.

Wakefield, A. (2003). *Selling Security. The Private Policing of Public Space.* Cullompton: Willan Publishing.

Walker-Munro, B. (2023). Hyper-collection: A possible new paradigm in modern surveillance. *Surveillance & Society*, 21(2), 120 – 138.

Zedner, L. (2006). Policing before and after the police. *British Journal of Criminology*, 46, 78 – 96.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power.* New York: Profile Books PublicAffairs.

Rose Lunde and Peter Scharff Smith

45 Punishment

Abstract: A key question when it comes to contemporary and future practices of punishment, is how these are influenced by the continuous rise of digital technologies, and in turn, how this might affect the purpose, as well as the function of punishment. This chapter addresses digitalization of punishment in connection with three themes: first, surveillance and control, where we will zoom in on the use of electronic monitoring, which has been employed in many jurisdictions. Secondly, we will look at prisons in digitized societies, focusing on video visits, digital surveillance in prisons, internet access, and the recent introduction of smart prisons. Thirdly, we will discuss the development and possible future of punishment in light of the continuous development of digital societies.

Keywords: prisons, punishment, surveillance, electronic monitoring

Introduction

A key question when it comes to contemporary and future practices of punishment, is how these are influenced by the continuous rise of digital technologies, and in turn, how this might affect the purpose, as well as the function of punishment. How will the relationship between punishment and society develop with increasing digitization?

The purpose of punishment has changed over time and varies from one jurisdiction to another, but nevertheless tends to focus on a combination of prevention (deterrence, rehabilitation, and incapacitation) and revenge. The latter motive is often portrayed as retrospective and moralistic, while crime prevention is typically considered utilitarian and forward looking. However, having moral aims and purposes of punishment can also be understood as a way of creating justice, thereby strengthening social cohesion in a given society. Various systems of punishment have prioritized these aims very differently over time, sometimes focusing on treatment and rehabilitation of offenders, sometimes on a combination of general deterrence, incapacitation, and revenge. If we look at the execution of punishment, for example in the form of imprisonment, further motives and considerations often enter the picture. These include security concerns (maintaining order and avoiding escapes), punitive concerns (factors that strengthen the element of punishment) as well as the question of how to respect prisoners' rights. This last point is often approached through the principle of normalization and aims to create prison conditions that resemble the outside world (Engbo and Smith, 2012).

If we look at all these various purposes, motives, and concerns from the perspective of digitalization (see Digital by Wernimont), there is little doubt that security concerns have been some of those most readily open to technological advances. Digital security and surveillance technologies have quickly been incorporated into the

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-046

infrastructure of prisons and penal systems to keep prisoners and pre-trial detainees under control. In contrast, digital technologies have been employed much more reluctantly for rehabilitative purposes. Perhaps the most dramatic illustration of the very selective use of digital technology in prisons is the fact that the inhabitants of these institutions are typically still cut off from the internet, even though people on the outside rely increasingly on continuous internet access in connection with almost all aspects of their daily lives.

This chapter will be structured according to three main themes. First, we will discuss digitalization of punishment in connection with surveillance and control. Here we will zoom in on the use of electronic monitoring, which has been employed in many jurisdictions. Secondly, we will look at prisons in digitized societies, focusing on video visits, digital surveillance in prisons, internet access, and the recent introduction of smart prisons (Kaun and Stiernstedt, 2020). Thirdly, we will discuss the development and possible future of punishment in light of the continuous development of digital societies.

Electronic monitoring

Global positioning (GPS) and radio frequency (RFID) systems have a diverse range of uses, such as determining the whereabouts of persons or objects through tracking of location data. Location tracking is used in the detection, investigation, and punishment of crime, the latter of which is done through electronic monitoring (EM). EM is the practice of holding convicted or remanded individuals under surveillance outside of prison, using GPS or RFID technology, generally in the form of a 'shackle' worn on the ankle (Nellis, 2021). Offenders under EM may continue living in their homes, but can often only leave to attend work, medical appointments, or to meet a parole officer. Alcohol monitoring devices that deliver data on consumption in real time are sometimes implemented in EM devices (Nellis, 2021). EM is increasingly used in place of prison for non-violent crimes; however, the technology is also used for monitoring of persons on remand (Lehman, 2022). Additionally, victims of intimate partner violence can be given GPS-enabled 'panic button' devices that ring directly to police if triggered; 'reverse EM' devices are also used to monitor persons under restraining orders (Daems, 2020). EM can also be used to enforce restrictions on offenders granted conditional release, such as by monitoring the whereabouts of persons convicted of sex offenses (Daems, 2020).

Proponents of EM argue for its potential as an alternative punishment that reduces the harms associated with prison by allowing offenders to serve their sentence at home, maintain their occupation, and have contact with family. Hence, EM can reduce issues with overcrowding, allow sentences to be carried out at reduced cost, and separate out offenders deemed unlikely to reoffend from those convicted of more serious offenses. The main argument in favor of EM, aside from the potential economic benefit, is that it removes the institution of prison from the equation entirely, thus lessening

the associated negative effects such as stigma, increased recidivism risk, isolation, addiction, and poor mental and physical health outcomes (Kaylor, 2022; Lehman, 2022).

Opponents of EM argue that many of the problems associated with prison are also present when the 'walls' are removed, and that EM instead represents an extreme form of deprivation of liberty, 'prisonizing' the homes, private lives, and even bodies of offenders (Campello, 2023). Family members can experience EM as a kind of punishment by proxy, and EM may impinge on the right to private and family life (Lehman, 2022). In some jurisdictions, costs associated with the device are borne by the wearer, in addition to needing to keep it charged at all times (Kirk, 2021). Devices that cannot be covered or which emit noise may subject the wearer to stigma; being under constant surveillance also psychologically taxing, in addition to the physical inconveniences of EM devices (Kilgore and Dolinar, 2023). Whilst the wearer appears to be part of free society, their reality is constricted and can be socially isolating. From this perspective, EM may be as invasive and harmful as prison, as the surveillance and control mechanisms of the criminal justice system move into the private sphere. Lastly, it is still unclear whether EM improves recidivism rates compared to prison, and the increase in its use has led to concerns around net widening and increased overall surveillance (Kirk, 2021; Kilgore and Dolinar, 2023). The technology also raises questions of data privacy, as well as public safety, if trends towards increasing use of EM continue in the future.

Prisons in digitized societies

Despite the extensive use of fines, as well as alternative punishments such as EM, community service, and mediation, the use of physical incarceration in prisons remain a core element of any modern system of punishment. The prison is a centuries-old institution which has proven notoriously difficult to reform in any fundamental way since the breakthrough of the modern penitentiary during the late 18th and early 19th centuries (Smith, 2004). Despite significant changes in the surrounding societies since that time, prison life has arguably remained relatively stable in its content and character. To explore the extent to which digital technologies have had an impact on prison life in recent decades, we will in the following zoom in on video visits; surveillance in prisons; and the question of internet access in these institutions.

Video visits

Several jurisdictions include video-visits as part of their institutional setup in all or some of their prisons. The pandemic was without any doubt an enabler in that regard even in nations well known to be especially tough on crime, such as the USA (Muñiz et al., 2024: 92). As in-person visits were suspended due to the risk of contamination, many prisons and correctional services in different parts of the world began introducing

video-visits (Lundeberg and Smith, 2022; Muñiz et al., 2024). In the USA towards the end of 2021, remarkably, more than half of all states (65.7 %) offered video-visits to all their prisoners and an additional six states (11.5 %) offered video visits to some of their residents (Muñiz et al., 2024: 91). In Norway, the Norwegian correctional service also introduced video-visits on a large scale in 2020 through the use of more than 800 iPads (Lundeberg and Smith, 2022).

However, the degree to which these technologies are actually available on the ground in prisons and how they are used varies greatly. Even in Norway, in a situation where the number of iPads amounted to around one quarter of the entire prison population, access was not assured. In the first phase of the pandemic with severe lockdowns, 30 % of prisoners reported that they had not received any offer of video visitation (Lundeberg and Smith, 2022). Additionally, many of those who used the iPads experienced significant technical problems with the actual execution of video visits (Lundeberg and Smith, 2022); similar problems have been experienced by prisoners in other jurisdictions (Murdoch and King, 2020).

Additionally, the reasons for using video visits (especially before and after the pandemic) vary significantly. In the US, for example, some correctional authorities have introduced video visits primarily to reduce costs associated with in-person visitation (Murdoch and King, 2020). A different logic has to do with providing prisoners with enhanced opportunities to "maintain family relationships and engage in programs aimed at optimizing their post-release circumstances and rehabilitation" (McKay, 2022: 100). Along these lines we see recent research studying the possible effects of prison video visits on recidivism (Duwe and McNeeley, 2021).

A third line of reasoning is to implement digital technologies including video-visits not primarily to rehabilitate and fight crime, but as a way of strengthening the rights of prisoners (Smith, 2013). In other words, there is a vast difference between the rationalities behind video visits for prisoners, which will undoubtedly be reflected in future policies and practices in this area.

Digital surveillance in prisons

Surveillance and control are essential parts of the fabric of prisons. Surveillance provides intelligence to prison staff about the whereabouts, actions, and possible intentions, of prisoners at any given time. In addition to gathering intelligence and recording evidence of potential criminal activity (see Intelligence by Gundhus and Lundgaard), the concept of surveillance is also used to maintain security and enforce rules. This is illustrated by Bentham's (1791) panopticon model, which is still in use today, designed to give prisoners the sensation of potentially being under observation at all times. Guard towers and cameras can both allow for just a single officer to collect data and maintain control through observation. Relative to other societal institutions, prisons have generally been reluctant to spend resources on adapting digital technologies, except for those applicable to surveillance and control. Examples of these tech-

nologies are improved security, camera and alarm systems, and x-ray scanners for people and packages entering the prison. Said technologies tend to be most attractive if they also promise improved efficiency and reduced expenditure: for example, installing more cameras may theoretically lead to reduced staff numbers.

In the crux of both control and concerns for prisoners' welfare, everyday objects are often adapted for use in prison to prevent them being used as weapons or as a means of suicide (Hughes and Metzner, 2015). Technologies specifically aimed at preventing harm and suicide have been developed for installation in prison cells, such as electronic respiration and motion detection via RFID (hereafter ERMD), which alert prison officers to 'unusual' activity (Hayes, 2013). In the case of Norway, ERMD has been approved for use in police custody as well as in solitary confinement cells in prisons, which are often used as both disciplinary measures and to isolate prisoners under suicide watch (Kriminalomsorgensdirektoratet, 2021). How this will impact current prison practices remains to be seen; this technology also raises questions about both the efficacy (relative to other surveillance methods and in terms of harm prevention) and ethics of its use.

Internet access

The arguably most pronounced illustration of the very selective use of digital technology in correctional settings is the fact that prisoners are typically completely cut off from the internet, whilst people on the outside rely increasingly on internet access in connection with their daily lives. Normally, when discussing internet use there is talk of a digital gap in the sense that the older generations have been lagging behind the younger generations and that some parts of the world have been lagging behind other, more developed parts of the world (Jørgensen and Smith, 2012). However, we find an even more dramatic digital gap when comparing incarcerated people with those living in the free world (Smith, 2013).

According to statistics from ITU three-quarters of the world's population (10 years or older) owned a mobile phone in 2022 (ITU, 2022: 17), and two-thirds used the internet (ITU, 2022: 1). In Europe and the Americas, the figures were 89 % and 83 % internet users respectively (ITU, 2022: 2). Additionally, in Europe 98 % of the 15- to 24-year-olds use the internet (ITU, 2022: 2). Remarkably, on a global level there are now significantly more mobile cellular telephone subscriptions than inhabitants (ITU, 2022: 8). Furthermore, in recent years the growth of mobile broadband subscriptions has been explosive, so that there are now 87 such subscriptions per 100 inhabitants in the world, and 110 and 113 in Europe and the Americas respectively (ITU, 2022: 10).

In stark contrast to these statistics, prisoners can typically spend years on end without any kind of internet access. In some jurisdictions, some level of access is allowed in the more open regimes (internet cafes in open Danish prisons for example), although these remain an exception to the rule. However, the development of so-called 'smart' prisons might change this.

Smart prisons

The term 'smart' is used widely to describe networked technologies that can communicate with each other, automate and adapt to tasks, and which have an increased functionality through doing so; including objects that utilize these technologies such as buildings, environments, and devices (Kaun and Stiernstedt, 2020). A related concept is that of *datafication* (see Datafication by Chan), whereby most aspects of life are captured electronically and turned into data which happens in conjunction with mass surveillance and the extraction and creation of large amounts of user data (Sadowski, 2019). The notion of 'smart' technologies in prisons was first mentioned in 2005 to refer to Finnish prison cells containing motion sensor technology (Murphy, 2005); however, prisons were until the 2010s not generally subject to datafication. Even in the 2020s, prisons typically do not enable access to mobile telephones, internet, or streaming services. Yet, many prisons built around this time advertise themselves as 'smart,' the concept of smart prisons has thus become of interest to researchers. Policymakers have used the etymology of being 'smart on crime' to refer to less punitive punishment policies (Altheide and Coyle, 2006); and we can draw parallels with how similar language is used to justify prisons' use of smart technologies.

What these institutions seem to have in common is their purposeful use of technology incorporating various aims and purposes of punishment. These technologies are often advertised explicitly as being used for rehabilitative purposes and to improve the prison environment for both prisons and staff. This could include online booking systems for health and welfare services, videoconferencing, digital access cards to cells (which replace the authoritative rattle of officers' keys), and the training of prisoners to operate in digital societies after release, such as assisting with online banking and setting up digital identification. As stated earlier, access to internet is still a contested issue. Some smart prisons allow access to heavily monitored 'intranets' rather than unfettered access to the global internet (Robberechts and Beyens, 2020). Adoption of smart technologies in prisons appears to have been spurred on by the COVID-19 pandemic, as with video visitation mentioned earlier. However, many prison technologies can also be seen as invasive and an extension of carceral control, while creating greater physical distance between prison staff and the persons in their care (McKay, 2022). Many prisons are built panopticon-style, with centralized guard posts that allow one or few officers to observe and facilitate activity (such as opening doors) remotely. Scanning of objects and bodies entering the prison, improved drug testing, GPS and RFID for both location, motion, and respiration detection, high-definition cameras, security systems, and cloud storage of sensitive information are some of the ways in which control, security, surveillance, and deterrence make themselves known within the technological fabric of smart prisons. Theoretically, there are few limits to how punitive and controlling (or alternatively rehabilitative and humane) a prison could become with the assistance of technology. A valuable avenue for research is to assess how smart technologies impact prison practices and prisoners' experiences of incarceration in

the future; much uncertainty hinges upon how the human aspects of prison life and interpersonal relationships between prisoners and staff, often summed up as *dynamic security*, will be affected as *static security* (the physical elements of security and surveillance, such as cameras, alarms, and walls) (Smith, 2022) becomes more automated (see Automation by Mann).

Possible future scenarios

Broadly speaking, one can envisage different future scenarios concerning the potential uses of, and approaches to, digital technologies within the realm of punishment:

- A high-tech revival of the abolitionist movement, focusing on the principle of normalization. Even prisoners are always online and can reach new heights of communicative and online freedoms, whilst physically restricted.
- Technology is weaponized for punitive purposes and increased security, minimizing contact with the external world. Prisoners are only able to access 'prison cloud' intranets, and prison staff communicate and operate solely online, further increasing isolation within prisons.
- Increased diversion away from prisons through widespread adaptation of EM and other technologies, reserving prison space for those deemed most dangerous or in most need of help.

In summary, digital technologies have the potential to fundamentally rearrange the landscape of punishment. This calls for increased empirical and theoretical attention from the research community as digital societies continue to evolve.

Suggested reading

Daems, T. (2020). Electronic monitoring in a culture of surveillance. In *Electronic Monitoring: Tagging Offenders in a Culture of Surveillance* (pp. 1 – 22). Cham: Palgrave Pivot.

Kaun, A., & Stiernstedt, F. (2020). Doing time, the smart way? Temporalities of the smart prison. *New Media & Society*, 22(9), 1580 – 1599.

McKay, C. (2022). The carceral automaton: Digital prisons and technologies of detention. *International Journal for Crime, Justice and Social Democracy*, 11(1), 100 – 119.

Smith, P. S. (2013). Imprisonment and internet-access – human rights, the principle of normalization and the question of prisoners' access to digital communications technology. *Nordic Journal of Human Rights*, 30(4), 454 – 482.

References

Altheide, D. L., & Coyle, M. J. (2006). Smart on crime: The new language for prisoner release. *Crime, Media, Culture*, 2(3), 286 – 303.

Bentham, J. (1791). *Panopticon: Postscript: Containing a Plan of Management for a Panopticon Penitentiary-House.* London: Payne, London.

Campello, R.U. (2023). Human-machine interface. In R. U. Campello (ed.), *Short Circuit: Electronic Monitoring and the Crisis of the Brazilian Prison System.* Cham: Springer.

Daems, T. (2020). *Electronic Monitoring.* Cham: Palgrave Pivot.

Duwe, G., & McNeeley, S. (2021). Just as good as the real thing? The effects of prison video visitation on recidivism. *Crime & Delinquency*, 67(4), 475 – 497.

Engbo, H. J., & Smith, P. S. (2012). *Fængsler og menneskerettigheder.* Copenhagen: Djøf forlag.

Hayes, L. M. (2013). Suicide prevention in correctional facilities: Reflections and next steps. *International Journal of Law and Psychiatry*, 36(3 – 4), 188 – 194.

Hughes, K. C., & Metzner, J. L. (2015). Suicide risk management. In R. L. Trestman, K. L. Appelbaum, & J. L. Metzner (eds.), *Oxford Textbook of Correctional Psychiatry* (pp. 237 – 244). New York: Oxford University Press

ITU. (2022). *Measuring Digital Development. Facts and Figures 2022.* ITU Publications. https://www.itu.int/itu-d/reports/statistics/facts-figures-2022/

Jørgensen, R. F., & Smith, P. S. (2012). Fængsler og internet – bør frihedsberøvede have adgang til nettet? *Social Kritik*, no. 132.

Kaylor, L. E. (2022). Electronic monitoring as an alternative to incarceration as part of criminal justice reform. In E. Jeglic & C. Calkins (eds.), *Handbook of Issues in Criminal Justice Reform in the United States* (pp. 701 – 716). Cham: Springer.

Kaun, A., & Stiernstedt, F. (2020). Doing time, the smart way? Temporalities of the smart prison. *New Media & Society*, 22(9), 1580 – 1599.

Kilgore, J., & Dolinar, B. (2023). Cages without bars are widening the net: The explosion of electronic monitoring. *Prison Legal News*, 34, 1.

Kirk, G. (2021). The limits of expectations and the minimization of collateral consequences: The experience of electronic home monitoring. *Social Problems*, 68(3), 642 – 657.

Kriminalomsorgensdirektoratet. (2021). *Innføring av pust og bevegelsessensor i Kriminalomsorgen.* Rundskriv KDI 2/2021.

Lehman, P. (2022). The other stay-at-home order: Electronic monitoring and the employability of criminal defendants after Covid. *Washington University Jurisprudence Review*, 15(1), 171 – 202.

Lundeberg, I., & Smith, P. S. (2022). Straff og etikk i pandemiens tegn – Dugnad og samfundssind i fengslene? In T. Fredwall & G. Heivoll (eds.), *Straffens etikk* (pp. 293 – 326). Oslo: Scandinavian Academic Press.

McKay, C. (2022). The carceral automaton: Digital prisons and technologies of detention. *International Journal for Crime, Justice and Social Democracy*, 11(1), 100 – 119.

Muñiz, J. O., Corcoran, F., Marzougui, J., Shlafer, R., Eddy, J. M., & Dallaire, D. (2024). Towards family preservation: A systematic jurisdiction analysis of prison visitation policies during the COVID-19 pandemic. *Psychology, Public Policy, and Law 30*(1), 91 – 103.

Murdoch, D. J., & King, L. L. (2020). 'Not feeling like a caged animal': Prisoner perceptions of a remote video visitation system. *Journal of Crime and Justice*, 43(2), 212 – 227.

Murphy, M. (2005). Smart prison cells monitor inmates' safety. *Chemistry & Industry*, 10, 8.

Nellis, M. (2021). Towards predictivity? Immediacy and imminence in the electronic monitoring of offenders. In B. Arrigo & B. Sellers (eds.), *The Pre-Crime Society: Crime, Culture and Control in the Ultramodern Age* (pp. 341 – 364). Bristol: Bristol University Press.

Robberechts, J., & Beyens, K. (2020). PrisonCloud: The beating heart of the digital prison cell. In J. Turner & V. Knight (eds.), *The Prison Cell: Embodied and Everyday Spaces of Incarceration* (pp. 283 – 303). London: Palgrave Macmillan.

Sadowski, J. (2019). When data is capital: Datafication, accumulation and extraction. *Big Data & Society*, 6(1).

Smith, P. S. (2004). A religious technology of the self: Rationality and religion in the rise of the modern penitentiary. *Punishment & Society*, 6(2), 195 – 220.

Smith, P. S. (2013). Imprisonment and internet-access – human rights, the principle of normalization and the question of prisoners access to digital communications technology. *Nordic Journal of Human Rights*, 30(4), 454 – 482.

Smith, P. S. (2022). Dynamic security or corruption of authority? Normalization and prisoner–staff relations in Danish prisons. In B. Crewe, A. Goldsmith, & M. Halsey (eds.), *Power and Pain in the Modern Prison: The Society of Captives Revisited* (pp. 290 – 308). Oxford: Oxford University Press.≤

Jan Christoffer Andersen

46 Recruitment via social media

Abstract: This chapter explores the challenges and advantages of using social media to recruit individuals from hard-to-reach populations to participate in research. It provides reflections on methodology, ethics, communication, and positionality based on the authors' online recruitment and interactions with members of the incel (involuntary celibate) community.

Keywords: incels, social media, online interviews, ethics, reflexivity

Introduction

Incels hide their identity and their affiliation to the online incel community from their offline environments (Lounela and Murphy, 2023). They mistrust outsiders, such as researchers or journalists, who criticize them for promoting antifeminism, misogyny, male supremacy, violence, and domestic terrorism (O'Donnell and Shor, 2022). So, how do we reach them, and how do we talk to them? In this chapter, I will explore both challenges and advantages of using social media to recruit individuals from hard-to-reach populations to participate in research. I provide reflections on methodology, ethics, communication, and positionality based on my online recruitment and interactions with members of the incel community.

As criminologists, we encounter various challenges and limitations during online recruitment and must continuously assess our actions before, during, and after participant interactions. The first considerations arise before recruitment, including ethical questions of anonymity, privacy rights, informed consent, and the practicalities of safeguarding our research participants and storing the interview data safely. We also need to build trust and rapport with our participants for the recruitment process to be successful. Part of that is to create credibility as researchers by familiarizing ourselves with the technologies we use to establish and uphold contact, while rapport refers to forming a positive relationship that makes the interviewee comfortable enough to openly share their experiences and thoughts (Nalita and Busher, 2012). Thus, assessing how we communicate and present ourselves throughout recruitment is crucial while upholding professional boundaries when facing gatekeeping, suspicion, or propaganda. It is also vital to assess our safety when interacting with individuals who may have committed crimes. Although recruitment processes online do not necessarily include offline interactions, we can still be susceptible to harm, such as doxing (i. e., revealing our personal information online), harassment, threats, or hacking.

The challenges of using social media in the recruitment process are plentiful. However, there are several advantageous aspects as well. Social media recruitment offers logistic convenience, eliminates the need for travel or transportation, and provides access to audio recording tools on the computer during the interview. It also reduces fi-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-047

nancial costs and minimizes personal exposure when trying to reach hard-to-reach individuals. One significant benefit is the flexibility of interviews, though this can also be problematic if they stretch out for an extended period (see App-Based Textual Interviews by Bakken). Notably, online interactions allow participants to control what they wish to share about themselves and how to communicate with us. Below, I will reflect on my experience conducting research with members of the incel community and discuss some fundamental aspects to consider when preparing for and conducting recruitment via social media.

The online incel subculture

Incels, a portmanteau of involuntary and celibate, are predominantly heterosexual men who have created an online subculture around their inability to form sexual or romantic relationships with women. They share experiences and grievances of loneliness, social isolation, alienation, and romantic rejection (Hoffman et al., 2020). Incels believe they cannot change their 'inceldom' status due to women's perceived genetic selectivity, which leads them to reject incels as sexual and romantic partners in favor of better-looking men (Baele et al., 2021). Their online activity varies, with some incels posting more extensively than others, amplifying controversial voices within the incel subculture that can result in broad generalizations about the entire incel community (Baele et al., 2022). Despite this, the online subculture is antagonistic towards outsiders, espousing misogyny, antifeminism, and violent fantasies that separates its members from the out-group, which includes 'regular' people, women, and sex-havers (Nagle, 2017). Although acts of violence by incels are rare (Cottee, 2021), the group has been associated with several mass killings, which were committed in the US in 2014, Canada in 2018, and the UK in 2022 (Lounela and Murphy, 2023).

Methodology: Why online recruitment?

We can use open sources and social media on the internet to observe, gain access, and research the activity and content of specific hard-to-reach populations who engage in deviant or criminal activity, such as illegal drug dealing or gang activity (Bakken, 2021; Pyrooz et al., 2015). These sources can provide important insights into online communities, their members, and the spaces they frequent. In the case of incels, researching forums and social media communication is vital in understanding their worldview, subcultural logic, and identity creation (Andersen, 2023). The divide between life 'online' and 'offline' is not easily separated, as both are intrinsically linked (Fileborn, 2016). However, individuals can act differently online than offline, and direct interaction with our research participants is sometimes necessary to gain additional nuance and knowledge that enables qualitative, interpretative, and critical approaches (see Online Ethnography by Gibbs and Hall and Accessing Online Communities by Kaufmann).

Thus, interviewees can raise additional questions, provide relevant context, and address significant concerns that may not be apparent in secondary data, big data sampling, or surveys.

The internet and online media are crucial for incels to express, form, and maintain their subcultural identity. However, incels often frame their grievances with antifeminism and misogyny (Ging, 2019). Interviews can enable them to elaborate on why they identify as 'involuntary celibate' based on their life experiences and how they navigate the shame and stigma associated with adopting the term 'incel.' As part of my Ph.D. research project, I recruited fourteen individuals who identified as incels or former incels. I primarily used my personal social media account on Facebook to contact participants through private incel groups. I conducted semi-structured interviews regarding their life history using Zoom or Discord, depending on the platform participants felt most comfortable with. I kept my camera on throughout the interviews to build rapport and trust, fostering a positive relationship between us (Nalita and Busher, 2012). Ten participants openly shared personal information, such as their names, ages, and country of origin. However, four participants preferred to remain anonymous by turning their cameras off, with one even using a voice modifier to further conceal his identity.

The internet and social media have become essential aspects of our daily lives, facilitating the engagement and recruitment of research participants who were previously difficult to reach. In criminological research, combining online recruitment methods with personal interviews offers a valuable means of comprehensively understanding hard-to-reach populations and their experiences, particularly concerning potential involvement in marginalized, deviant, or criminal activities.

Ethics: Getting ready

Before recruitment begins, it is crucial to assess how we can maintain our well-being as researchers, as well as the well-being of our participants (Lumsden and Winter, 2014). This assessment involves identifying the demographics of interviewees and the online environments we intend to study to address our research questions accurately. We are also obligated to uphold our participants' rights to anonymity, confidentiality, privacy, and informed consent (Fossheim and Ingierd, 2015). While structuring our research, we must consider these concerns when applying for ethical approval from an ethics committee or review board. This process is essential for identifying potential problems that may arise through our recruitment methods.

The requirement for ethics approval varies depending on the institution or country (e. g., AoIR, 2019; BSA, 2017; NESH, 2019), and many of their guidelines problematize our use of covert and deceptive research methods when conducting online research. Some guidelines mandate transparency about our research identity when recruiting individuals online. It is, therefore, vital for us as researchers to consider how much personal information we are comfortable disclosing. First, we must assess our digital

profile and online footprints that could expose personal information we prefer to keep private. We then, if necessary, limit access to our private social media and any identifiable information about ourselves. It is also essential to adopt general online safety measures to prevent hacking and the misuse of our passwords by updating them regularly, enabling two-step authenticators, and using malware programs. As an exercise in online security and safety, I recommend partnering with other students or researchers to simulate a 'doxing' scenario, where participants attempt to gather as much information about each other as possible. This exercise aims to reveal any identifying or uncomfortable information that can maliciously be used against you online.

Participants can share sensitive, traumatic, and even incriminating information about themselves online (Sidoti, 2023). As criminologists, we sometimes work with individuals who have committed, are currently engaging in, or are planning to engage in criminal activities. Therefore, being informed about the ethical considerations and legal requirements concerning our obligations to notify the authorities of specific crimes is critical. However, we must also balance this with our participants' well-being and confidentiality by safeguarding our interviewees' anonymity and managing the data safely. An example of the tension between research confidentiality and legal demands emerged in the controversial Boston College Tapes case, where researchers were compelled to release interviews with former paramilitaries to authorities (Sampson, 2016).

Members of the incel community have hacked and publicly published the private information of people they want to shame, harass, and potentially harm (Nagle, 2017). Throughout my recruitment, I selected smaller incel groups on Facebook and minimized my online exposure to the broader incel community. These groups differed from larger, puritanical incel forums since they were private, allowed outsiders, and prohibited encouraging violence. Most of my participants revealed their identities; therefore, upholding their security, confidentiality, and privacy was crucial. They requested to remain anonymous due to concerns and fears about social stigma, governmental control, and backlash from other incels.

Access: Getting in touch

We can use snowball and convenience sampling to recruit individuals online (Baltar and Brunet, 2012). Establishing credibility, trust, and rapport when interacting with our subjects, however, remains essential. This process can be particularly time-consuming and challenging for individuals from hard-to-reach communities (Kaufmann and Tzanetakis, 2020). When we introduce ourselves, questions about our identity and research aims should be clear. However, tensions between us as researchers and our participants might still arise regarding power dynamics, conflicting agendas, and separating our public and private selves (Lavorgna and Sugiura, 2022b).

One approach is making ourselves known to the community by publishing recruitment posts within online spaces where wanted participants frequent. It should include

information about us as researchers, the research project, our professional contact information, and whom we want to interview. This approach casts a wide net as it is visible to active and inactive individuals within the community, which allows willing participants to ask further questions they might have. However, it can also expose us to potentially harmful community members, making it vital to determine if the online space is hostile towards researchers due to our positionality regarding gender, ethnicity, sexual orientation, or political affiliation. One measure to limit our exposure is by recruiting from smaller groups on social media, chat groups, or forums (see Researching Online Forums by Šupa).

Another approach involves contacting participants individually via direct messaging, text, apps, or e-mail (Bakken, 2022). Success varies depending on the participants' openness about their role in the online community and their trust in the conducted academic research. We must also stay flexible when contacting and recruiting potential participants since the instantaneous nature of online communication can pressure us to respond quickly (Fileborn, 2016). Some may only show initial interest, making us complete the interview on short notice, and might not respond later. Gatekeepers can play an essential role in gaining access to online recruitment and enable us access to forums, chat groups, or other relevant participants through recommendations or introductions (Banks, 2014). One way to create a positive relationship with gatekeepers online is by contacting them directly, establishing your role as a researcher, and expressing your interest in interviewing them due to their extensive knowledge of the community and their central role, such as being an administrator of a forum.

Snowball sampling was, in my case, relatively successful for the online recruitment of incels. Due to the anonymous and mistrusting nature of the incel community, few participants were close to or knew other incels personally. There were, however, exceptions, as some were part of smaller chat groups on Facebook or Discord servers that allowed them to communicate more openly. For example, one participant was unfamiliar and uncomfortable using Zoom as a communication tool. However, he instead wanted to communicate using Discord due to its familiarity, making it easier for him to stay anonymous. To accommodate this, I created a server to conduct the interview, allowing the participant to recruit and invite other incels he knew into the server after he completed his interview.

Communication: What to say?

Building trust and rapport with individuals online is an ongoing aspect of the recruitment and interview process. Criminologists often address sensitive topics and engage with vulnerable populations (see Vulnerability by Ranchordas and Beck). Online interactions can create flexible and informal social settings for participants, giving interviewees more control of the interview environment (Bakken, 2022). Although participants may not have total control, they can interrupt or pause interviews and choose how to communicate—through text, audio, or video—and what they wish to share

by anonymizing their identities using pseudonyms or encryption. However, this can also create power asymmetries and tension, especially when researchers are required to disclose their identities, unlike participants (Lavorgna and Sugiura, 2022a).

Online interaction reduces travel time and costs, enabling us to contact multiple people simultaneously. It can mediate the relationship of trust in various ways, where the anonymous and fleeting nature of online interactions can create a sense of distance between the researcher and the participant. However, the lack of physical presence can also make some participants feel more comfortable, with the possibility of testing our technical knowledge (Kaufmann and Tzanetakis, 2020), thereby increasing trust. Nevertheless, online research can present challenges when compared to face-to-face methods. Participants may quickly stop responding or withdraw from the study (Mardones-Bravo, 2023). For instance, when seeking informed consent from research participants, a challenge arises if interviewees withdraw after receiving an information letter and consent form, stop responding, or fail to attend scheduled interviews. The formality of the process can intimidate some—despite previous communication having been more relaxed and unproblematic. During my recruitment of incels, some stopped responding or blocked me at this stage. However, most were keen to share their experiences to correct misconceptions about their community and were less concerned about the consent form.

The written information letter given to participants should include information about the purpose of the project, the responsible institution, the reasons for the invitation, and what their participation entails (e. g., length of the interview and possible questions). It should also include a section about personal privacy, participants' rights, and contact information for further information—additionally, have a section about the possibility of withdrawing from the research project. Following the information letter, the consent form should include checkboxes for the participants to fill out and sign in written form. Nonetheless, obtaining written consent online can be challenging when interviewees have limited knowledge or access to programs that enable electronic signing. If it is impossible to get written consent, many ethical guidelines accept recorded oral consent as sufficient, provided the participant is fully informed.

Setting clear boundaries between our professional and personal lives is crucial in online research. Throughout my recruitment of incels, I used my personal Facebook profile, which blurred the lines between public and private (Andersen and Sugiura, 2024). I did this to build trust and rapport, demonstrating that I was not there to demonize incels and that I was not deceptively infiltrating their community. Most participants were more focused on conveying their perspectives than on me, but after some interviews, I received friend requests on my social media account. I politely declined these requests, explaining that I needed to maintain a professional distance to protect both the identity of my participants and my own privacy. I was concerned that my participants might react negatively to being turned down, potentially causing conflict (Lavorgna and Sugiura, 2022b). Fortunately, that was not the case in my experience.

Conclusion

Digital research offers unique opportunities and is essential for understanding the evolving landscape of internet-related harm. Online recruitment provides access to diverse and hard-to-reach populations, allowing for a more nuanced understanding of their behavior. The process, however, can be complex as our positionality, presence, and use of social media impact the situation. Nonetheless, we can manage these complexities with appropriate strategies and sensitivity to participants' rights:

- Ethical challenges: These include maintaining participants' anonymity and privacy rights, protecting sensitive information through proper data management, and obtaining informed consent.
- Methodological challenges: These involve addressing trust, safety, and communication issues on online platforms and comprehending power dynamics and personal boundaries in the context of digital recruitment.
- Security challenges: These require assessment and minimization of our digital profile and online footprints to avoid inadvertent exposure of our confidential information.

This chapter aims to inform and guide researchers and students about the intricacies of conducting recruitment via social media. Adopting a reflexive approach can help handle some of these concerns (Alvesson and Skoldberg, 2018). In other words, we must reflect on how our identity might influence participant recruitment and interactions throughout our research process online.

Suggested reading

Participant recruitment on social media:

- Fileborn, B. (2016). Participant recruitment in an online era: A reflection on ethics and identity. *Research Ethics*, 12(2), 97 – 115.
- Internet research ethics and data management:
- Fossheim, H., & Ingierd, H. (2015). *Internet Research Ethics.* Oslo: Cappelen Damm Akademisk.
- Reflexivity and positionality in criminological research and qualitative research:
- Lumsden, K., & Winter, A. (2014). *Reflexivity in Criminological Research: Experiences with the Powerful and the Powerless.* London: Palgrave Macmillan.
- Alvesson, M., & Skoldberg, K. (2018). *Reflexive Methodology: New Vistas for Qualitative Research.* Third edition. Los Angeles: Sage.
- Qualitative and online research in criminology:
- Faria, R., & Dodge, M. (2023) *Qualitative Research in Criminology: Cutting-Edge Methods.* Cham: Springer.

References

Alvesson, M., & Skoldberg, K. (2018). *Reflexive Methodology: New Vistas for Qualitative Research.* Third edition. Los Angeles: Sage.

Andersen, J. C. (2023). The symbolic boundary work of incels: Subcultural negotiation of meaning and identity online. *Deviant Behavior*, 44(7), 1081 – 1101. doi: 10.1080/01639625.2022.2142864.

Andersen, J. C., & Sugiura, L. (2024). Interacting with online deviant subcultures: Gendered experiences of interviewing incels. In R. S. Graham, S. G. Humer, C. S. Lee & V. Nagy (eds.), *The Routledge International Handbook of Online Deviance* (pp. 537 – 553). New York: Routledge. doi: 10.4324/9781003277675-35

AoIR. (2019). *Internet Research: Ethical Guidelines 3.0 Association of Internet Researchers.*

Baele, S. J., Brace, L., & Coan, T. G. (2021). From 'incel' to 'saint': Analyzing the violent worldview behind the 2018 Toronto attack. *Terrorism and Political Violence*, 33(8), 1667 – 1691. doi: 10.1080/09546553.2019.1638256.

Baele, S. J., Brace, L., Coan, T. G., & Naserian, E. (2022). Super- (and hyper-) posters on extremist forums. *Journal of Policing, Intelligence and Counter Terrorism*, 18(3), 243 – 281. doi: 10.1080/18335330.2022.2103386.

Bakken, S. A. (2021). Drug dealers gone digital: Using signalling theory to analyse criminal online personas and trust. *Global Crime*, 22(1), 51 – 73. doi: 10.1080/17440572.2020.1806826.

Bakken, S. A. (2022). App-based textual interviews: Interacting with younger generations in a digitalized social reality. *International Journal of Social Research Methodology*, 26(6), 631 – 644. doi: 10.1080/13645579.2022.2087351.

Baltar, F., & Brunet, I. (2012). Social research 2.0: Virtual snowball sampling method using Facebook. *Internet Research*, 22(1), 57 – 74. doi: 10.1108/10662241211199960.

Banks, J. (2014). Online gambling, advantage play, reflexivity and virtual ethnography. In K. Lumsden & K. Winter (eds.), *Reflexivity in Criminological Research* (pp. 289 – 300). New York: Palgrave Macmillan.

BSA. (2017). *Ethics Guidelines and Collated Resources for Digital Research: Statement of Ethical Practice Annex.* British Sociological Association.

Cottee, S. (2021). Incel (e)motives: Resentment, shame and revenge. *Studies in Conflict & Terrorism*, 44(2), 93 – 114. doi: 10.1080/1057610X.2020.1822589.

Faria, R., & Dodge, M. (eds.) (2023). *Qualitative Research in Criminology: Cutting-Edge Methods.* Cham: Springer.

Fileborn, B. (2016). Participant recruitment in an online era: A reflection on ethics and identity. *Research Ethics*, 12(2), 97 – 115. doi: 10.1177/1747016115604150.

Fossheim, H., & Ingierd, H. (eds.) (2015). *Internet Research Ethics.* Oslo: Cappelen Damm Akademisk.

Ging, D. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities*, 22(4), 638 – 657. doi: 10.1177/1097184X17706401.

Hoffman, B., Ware, J., and Shapiro, E. (2020). Assessing the threat of incel violence. *Studies in Conflict & Terrorism*, 43(7), 565 – 587. doi: 10.1080/1057610X.2020.1751459.

Kaufmann, M., & Tzanetakis, M. (2020). Doing Internet research with hard-to-reach communities: Methodological reflections on gaining meaningful access. *Qualitative Research*, 20(6), 927 – 944. doi: 10.1177/1468794120904898.

Lavorgna, A., & Sugiura, L. 2022a. Blurring boundaries: Negotiating researchers' positionality and identities in digital qualitative research. *Italian Sociological Review*, 12(7S), 709 – 727. doi: 10.13136/ISR.V12I7S.578.

Lavorgna, A., & Sugiura, L. (2022b). Direct contacts with potential interviewees when carrying out online ethnography on controversial and polarized topics: A loophole in ethics guidelines. *International Journal of Social Research Methodology*, 25(2), 261 – 267. doi: 10.1080/13645579.2020.1855719.

Lounela, E., & Murphy, S. (2023). Incel violence and victimhood: Negotiating inceldom in online discussions of the Plymouth shooting. *Terrorism and Political Violence*, 36(3), 344 – 365. doi: 10.1080/09546553.2022.2157267.

Lumsden, K., & Winter, A. (eds.) (2014). *Reflexivity in Criminological Research: Experiences with the Powerful and the Powerless.* London: Palgrave Macmillan.

Mardones-Bravo, D. (2023). Online methods in qualitative criminology. In R. Faria & M. Dodge (eds.), *Qualitative Research in Criminology* (pp. 85 – 102). Cham: Springer.

Nagle, A. (2017). *Kill All Normies.* Winchester: Zero Books.

Nalita, J., & Busher, H. (2012). Internet interviewing. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (eds.), *The SAGE Handbook of Interview Research: The Complexity of the Craft* (pp. 177 – 188). Thousand Oaks, CA: SAGE Publications.

NESH. (2019). *A Guide to Internet Research.* National Research Ethics Committees.

O'Donnell, C., & Shor, E. (2022). 'This is a political movement, friend': Why 'incels' support violence. *The British Journal of Sociology*, 73(2), 336 – 351. doi: 10.1111/1468 – 4446.12923.

Pyrooz, D. C., Decker, S. H., & Moule, R. K. (2015). Criminal and routine activities in online settings: Gangs, offenders, and the internet. *Justice Quarterly*, 32(3), 471 – 499. doi: 10.1080/07418825.2013.778326.

Sampson, F. (2016). 'Whatever you say… ': The case of the Boston College tapes and how confidentiality agreements cannot put relevant data beyond the reach of criminal investigation. *Policing*, 10(3), 222 – 231. doi: 10.1093/police/pav034.

Sidoti, C. (2023). Smart researching in criminology: Virtual ethnography at the edge. In R. *Faria & M. Dodge (eds.), Qualitative Research in Criminology* (pp. 53 – 67). Cham: Springer.

Maryia Šupa

47 Researching online forums

Abstract: The chapter focuses on online forum research in criminology. It discusses existing methodologies and theoretical challenges faced by researchers, and future developments in the area. As socio-technical systems, online hacking forums provide a valuable source of information about the nature of illicit economies and construction of identities and norms in digital societies.

Keywords: online offending, online forums, research methodologies

Online forums are discussion platforms with multiple contributions from online participants, relevant as sources of observational and interactive data for criminologists. They differ from other online services by providing a space for community-building and long-term collection of knowledge that is structured into searchable, easily accessible archives. Many forums exist as grassroots initiatives independent of corporate online service providers such as Facebook or Google, maintained and moderated by volunteers and administered in accordance with community rules. Online forums have developed as modernized web versions of a previous generation of discussion tools such as bulletin boards and newsgroups. As a platform of computer-mediated communication, forums are a hybrid techno-social tool that combines a technological infrastructure with authentic social interactions (see Infrastructures by Grisot and Parmiggiani).

Online forum research is a subset of the broader research on online communities. It started in tandem with the mass adoption of the web in the late 1990s and early 2000s. Disciplines as diverse as anthropology, education, folklore studies, gender studies, healthcare, psychology, sociology, and, last but not least, criminology included online phenomena in their research. It is a versatile and flexible research field offering unique practical insights. It differs from other online research object such as news websites, social networking services, or online shopping platforms. Online forum research provides grounds for interdisciplinary collaboration between different disciplines, methodologies, and topic areas. A single forum may yield data useful for a range of research questions and analytical approaches.

In criminology, online forums provide the opportunity to reach under-researched and hard-to-access groups such as online offenders, participants in illicit online trade, or stigmatized crime survivors (see Accessing Online Communities by Kaufmann). It facilitates the research of sensitive or covert topics such as sexual abuse. Via forums, researchers can reach transnational and intercultural communities from all over the world. They gain access to both openly published content and hidden content (see Darknet by Tzanetakis), thus facing a host of ethical dilemmas.

Online forums provide access to at least four types of communities that are relevant to digital criminology:

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-048

- Communities discussing illicit activities that rely on computers and networks, such as distribution of mobile malware (Grisham et al., 2016), sexually explicit scamming (Pastrana et al., 2019), and stolen data sales (Hutchings and Holt, 2015).
- Communities discussing other clearly illicit activities that also occur offline, including illicit drug sales (Ferguson, 2017), paedophilia (Holt et al., 2010), or illegal consumption of copyrighted material such as music (Caldwell Brown, 2016).
- Communities discussing deviant, although not necessarily illicit activities, such as self-injury (Rodham et al., 2016) or incel subculture (Liggett O'Malley and Helm, 2022).
- Stigmatized communities, including support groups for victims of specific crimes, such as rape survivors (O'Neill, 2018).

The main theoretical approaches used by online forum researchers reflect the paradigmatic divide between positivist and critical research agendas. Positivist approaches are most often represented by rational choice theory (e. g., Smirnova and Holt, 2017; Holt et al., 2015) and situational crime prevention (Chavez and Bichler, 2019), while critical approaches, although much less common, include cultural criminology (Alashti et al., 2022). Criminological studies of online forums frequently lack a thorough discussion of how their theoretical approaches have been adapted to an online context, or how online context expands or transforms key theoretical concepts. For instance, while online hacking forums are often identified as subcultures, scant argumentation is provided about their subcultural traits, communal values, or the changes that occur when members of a subculture interact online.

Methodologies

Online forums have been researched using different methodologies, including quantitative, qualitative, and mixed methods. Quantitative approaches most often rely on large datasets such as the Cambridge Cybercrime Centre's 'CrimeBB' database of online hacking forum posts (Pastrana et al., 2018) or other automatically collected (Munksgaard and Demant, 2016) or leaked (Holt and Dupont, 2019; Overdorf et al., 2018) forum databases. Natural language processing (NLP) and machine learning (ML) are used to analyze large datasets. They combine multiple steps and techniques of analysis. Results allow researchers to determine interests and developmental pathways of key actors in online hacking forums (Hughes et al., 2019), to describe specific offenses (Pastrana et al., 2019), and to predict private interactions between forum members based on publicly posted data (Overdorf et al., 2018). Topic modeling has been used for simpler tasks, such as describing the key topics of online forums specializing in payment card scams (Kigerl, 2018), or analyzing and classifying political discourses on cryptomarket forums (Munksgaard and Demant, 2016). Sentiment analysis has been utilized for models predicting the onset of online offences (Deb et al., 2018). Social network analysis has been used to determine the overall social structures of forum communi-

ties (Pete et al., 2020) and characteristics of sub-communities (Garg et al., 2015). Content analysis, grounded in statistics, has also been conducted. An example is the analysis of variables associated with individuals that have requested membership in a closed hacking forum (Holt and Dupont, 2019).

Qualitative approaches may be carried out on purposive samples from the same large datasets. For example, based on data from the CrimeBB database, researchers have conducted a thematic content analysis of offenses targeting the Internet of Things (Bada and Pete, 2020). Custom data may also be collected specifically for the project. Qualitative content analysis of threads and messages posted in online forums has been used for the analysis of information security advice (Chavez and Bichler, 2019), administration of malware (Hutchings and Clayton, 2017) or development of pedophile subcultures (Holt et al., 2010). Online ethnography and participant observations are long-term projects that add interactions with forum members to content analysis. They have been conducted to explain, for instance, online drug trade (Ferguson, 2017).

Mixed methods studies have combined analysis of data from online forums with other sources of data, e. g., interviews (Collier et al., 2021; Alashti et al., 2022), integrated online and offline ethnography (Potter, 2017; Ferguson, 2017), or incorporated quantitative and qualitative methods for analyzing the same data (Bada et al., 2021).

Contributions

A key contribution of online forum research to criminology is the debunking of popular myths about the internet and online offending. One such myth is that communities labeled as online hacking forums support intense illicit activity (see Hacking by Wall). On the contrary, studies have found the majority of content on these forums is unrelated to offending (Siu et al., 2021), or that the majority of the users in online hacking forums posed little actual risk (Holt et al., 2014). Another myth is that communities focused on offending prevail on non-public networks such as Tor (referred to popularly as the darknet). On the contrary, there are many publicly accessible spots on the web where illicit activity occurs including on major platforms like Facebook. There are also specific myths about the mechanisms of the illicit economy, e. g., the myth that anonymity is the key advantage of cryptocurrency in illicit trade (Butler, 2021).

Beyond debunking myths, criminological research of online hacking forums sparks discussions about the broader implications of the illicit economy facilitated by technology (Holt, 2012). One of such frames interprets participation in illicit online activity as an act of resistance to mainstream capitalism (Maddox, 2020). Analysis of online forums also raises questions about the late modern condition, such as issues of a fragmented, mediated postmodern identity apparent on hacking forums. Issues of trust, risk, uncertainty (Yip et al., 2013), or boredom (Collier et al., 2021) may also be connected to global and local social structures, experienced by offenders and non-offenders alike.

There are several benefits of using online forums as a data source. Their anonymous or pseudonymous nature and open nature of discussions provide convenient access to data on a wide variety of topics, including sensitive and taboo topics. The practical value of positivist online forum research for criminal justice and law enforcement lies in the definition of crime scripts (Hutchings and Holt, 2015; Chavez and Bichler, 2019), intelligence, such as identification of illicit supply chains (Bhalerao et al., 2019) or new malware (Grisham et al., 2017), risk modeling and prediction (Deb et al., 2018), and prevention tactics. Since online forums are communities, they represent distinct cultural milieus. Understanding them is critical to practical outcomes of critical research, creating insight into how norms are formed, evaluated, deliberated, and breached. From the point of view of critical theory, it is also possible to observe how external discourses find their expression in online forums and reflect the broader configurations of discursive power.

Forums are not just content publishing platforms but communities that bring together active contributors and a network of committed members. They allow exploring how community members create and deliberate the content, how they express the rationale behind different standpoints, and how they maintain their social ties. Forums are instrumental for the construction and performance of a collective identity (Maltby et al., 2018: 1775) by the community. Their structures facilitate the building of trust, which has special significance for communities focused on offending (Yip et al., 2013: 535). Other community functions include control, coordination, social networking, and management of uncertainties and risks (Yip et al., 2013). All of these help understanding of how norms and transgressions are negotiated and evaluated.

Challenges

Online forum research also presents practical challenges. One challenge, especially for quantitative analysis of online forum data, is its informal and unstructured language. If data is collected automatically, additional tools are needed to filter out spam and unsuitable data. Online forum posts are often short, weakly structured, include non-conventional language, emoticons, and abbreviations that make structural and algorithmic analysis of the text difficult (Ferguson, 2017: 694).

Researchers face a host of ethical dilemmas (see Ethics by Markham). They must delicately balance between disclosing enough information about their research process to make it reproducible, and preserving the anonymity of the studied communities, especially if they are vulnerable or non-public (franzke et al., 2020). Some online forum content may be extreme and harmful to the researcher and requires additional scrutiny and preparation, as well as a situated approach to ethics (Kelley and Weaver, 2020). It is often impossible to request consent from online forum members, and much of the research is carried out covertly, which leads to ethical challenges. While researchers benefit from publishing their findings, there are few opportunities for forum users to familiarize themselves with the research, react to it, or share some of the possible

benefits if research is done covertly. Online forum contents were not initially created as research data, so further questions arise about the ethics of using them.

Beyond researching forums

Online forums, collections of topical posts and discussion threads interactively produced by their users, are not the only platforms where discussions of potentially illicit activities occur. Other platforms should not be overlooked when considering potential data sources. These may include platforms such as Reddit which hosts community discussions, but is less structured than online forums, and provides additional tools such as upvoting or downvoting posts. Applications for chatting and group communication such as Telegram, Signal, or Discord may be accessed. However, their content is unstructured and may quickly disappear, necessitating approaches that allow description, interaction, and analysis as an ongoing process rather than collecting data in a single step. Illicit online marketplaces may provide valuable criminological insight, and should not be confused with online forums in general. While some marketplaces exist within forum ecosystems, the most prominent ones are exclusively focused on trade and provide additional functionality such as escrow services—the payment is held by the marketplace until the seller and buyer complete their transaction.

Summary of key points

- Online forum research is a subset of online community research focusing on online forums—discussion platforms which allow communities to create, deliberate, and organize various forms of knowledge, and to create and maintain social ties via online interactions. They are valued as a source of scientific data for the variety of topics that may be addressed and large diversity of possible approaches.
- Online forum research is significant for digital criminology as a source of knowledge about communities discussing online and offline illicit activities, as well as deviant and stigmatized groups. Although currently positivist theoretical frameworks prevail in criminological works about online forums, critical approaches also exist.
- Criminological research of online forums facilitates the debunking of several popular myths about online and offline offending. It encourages broader discussion about the nature of illicit economies and construction of identities and norms in digital societies.
- Different quantitative, qualitative, and mixed methods designs have been applied in online forum research. They range from automated analysis based on Natural Language Processing of large data samples from multiple forums to in-depth thematic coding by hand of purposively sampled messages. While much research in this area is based on content only, participant online ethnography allows to lever-

age the direct involvement of a researcher with an online forum community (see Online Ethnography by Gibbs and Hall).
– The key benefits of online forum research are convenient access to published data and opportunity to observe how online communities function, including negotiation of norms and transgressions. The key challenges are ethical dilemmas, such as covert research or the use of data not produced for research purposes, linguistic complexities, and competing platforms that may be used to study similar communities.
– The boundary between online and offline life has become blurred and is continuing to do so. Most social practices, including illicit ones, become increasingly represented online, even if the activities themselves occur offline. Therefore, online forums will remain one of the significant sources of data about a large range of offending and deviant behavior (Potter, 2017: 3–4), including newly emerging forms of crime.

Suggested reading

franzke, a. s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* Available online at: https://aoir.org/reports/ethics3.pdf

Ferguson, R. H. (2017). Offline 'stranger' and online lurker: Methods for an ethnography of illicit transactions on the darknet. *Qualitative Research*, 17(6), 683–698.

Jemielniak, D. (2020). *Thick Big Data: Doing Digital Social Sciences.* Oxford: Oxford University Press.

References

Alashti, Z. F., Bojnordi, A. J. J., & Sani, S. M. S. (2022). Toward a carnivalesque analysis of hacking: A qualitative study of Iranian hackers. *Asian Journal of Social Science*, 50(2), 147–155.

Bada, M., & Pete, I. (2020). An exploration of the cybercrime ecosystem around Shodan. In *2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, 1–8.

Bada, M., Chua, Y. T., Collier, B., & Pete, I. (2021). Exploring masculinities and perceptions of gender in online cybercrime subcultures. In M.W. Kranenbarg & R. Leukfeldt (eds.), *Cybercrime in Context: The Human Factor in Victimization, Offending, and Policing* (pp. 237–257). Cham: Springer International Publishing.

Bhalerao, R., Aliapoulios, M., Shumailov, I., Afroz, S., & McCoy, D. (2019). Mapping the underground: Supervised discovery of cybercrime supply chains. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*, 1–16.

Butler, S. (2021). Cyber 9/11 will not take place: A user perspective of bitcoin and cryptocurrencies from underground and dark net forums. In: *STAST 2020: Socio-Technical Aspects in Security and Trust*, 135–153.

Caldwell Brown, S. (2016). Where do beliefs about Music Piracy come from and how are they shared? An ethnographic study. *International Journal of Cyber Criminology*, 10(1), 21–39.

Chavez, N., & Bichler, G. (2019). Guarding against cyber-trespass and theft: Routine precautions from the hacking community. *International Journal of Cyber Criminology*, 13(1), 101–116.

Collier, B., Clayton, R., Hutchings, A., & Thomas, D. (2021). Cybercrime is (often) boring: Infrastructure and alienation in a deviant subculture. *The British Journal of Criminology*, 61(5), 1407 – 1423.

Deb, A., Lerman, K., & Ferrara, E. (2018). Predicting cyber-events by leveraging hacker sentiment. *Information*, 9(11).

Ferguson, R. H. (2017). Offline 'stranger' and online lurker: Methods for an ethnography of illicit transactions on the darknet. *Qualitative Research*, 17(6), 683 – 698.

franzke, a. s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* Available online at: https://aoir.org/reports/ethics3.pdf

Garg, V., Afroz, S., Overdorf, R., & Greenstadt, R. (2015). Computer-supported cooperative crime. In R. Bohme & T. Okamoto (eds.), *Financial Cryptography and Data Security* (pp. 32 – 43). Heidelberg: Springer Berlin.

Grisham, J., Samtani, S., Patton, M., & Chen, H. (2017). Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 13 – 18.

Holt, T. J. (2012). Examining the forces shaping cybercrime markets online. *Social Science Computer Review*, 31(2), 165 – 177.

Holt, T. J., & Dupont, B. (2019). Exploring the factors associated with rejection from a closed cybercrime community. *International Journal of Offender Therapy and Comparative Criminology*, 63(8), 1127 – 1147.

Holt, T. J., Blevins, K. R., & Burkert, N. (2010). Considering the pedophile subculture online. *Sexual Abuse: A Journal of Research and Treatment*, 22(1), 3 – 24.

Holt, T. J., Smirnova, O., Strumsky, D., & Kilger, M. (2014). Case study: Advancing research on hackers through social network data. In G. E. Higgins & C. D. Marcum (eds.), *Social Networking as a Criminal Enterprise* (pp. 45 – 163). Boca Raton, FL: CRC Press.

Holt, T. J., Smirnova, O., Chua, Y. T., & Copes, H. (2015). Examining the risk reduction strategies of actors in online criminal markets. *Global Crime*, 16(2), 81 – 103.

Hughes, J., Collier, B., & Hutchings, A. (2019). From playing games to committing crimes: A multi-technique approach to predicting key actors on an online gaming forum. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*, 1 – 12.

Hutchings, A., & Clayton, R. (2017). Configuring Zeus: A case study of online crime target selection and knowledge transmission. In *2017 APWG Symposium on Electronic Crime Research (eCrime)*, 1 – 8.

Hutchings, A., & Holt, T. J. (2015). A crime script analysis of the online stolen data market. *British Journal of Criminology*, 55(3), 596 – 614.

Kelley, B., & Weaver, S. (2020). Researching people who (probably) hate you: When practicing "good" ethics means protecting yourself. *Computers and Composition*, 56, 1 – 9.

Kigerl, A. (2018). Profiling cybercriminals: Topic model clustering of carding forum member comment histories. *Social Science Computer Review*, 36(5), 591 – 609.

Maddox, A. (2020). Disrupting the ethnographic imaginarium: Challenges of immersion in the Silk Road cryptomarket community. *Journal of Digital Social Research*, 2(1), 20 – 38.

Maltby, S., Thornham, H., & Bennett, D. (2018). Beyond 'pseudonymity': The sociotechnical structure of online military forums. *New Media & Society*, 20(5), 1773 – 1791.

Munksgaard, R., & Demant, J. (2016). Mixing politics and crime – the prevalence and decline of political discourse on the cryptomarket. *International Journal of Drug Policy*, 35, 77 – 83.

Liggett O'Malley, R., & Helm, B. (2022). The role of perceived injustice and need for esteem on incel membership online. *Deviant Behavior*, 44(7), 1026 – 1043.

O'Neill, T. (2018). 'Today I speak': Exploring how victim-survivors use Reddit. *International Journal for Crime, Justice and Social Democracy*, 7(1), 44 – 59.

Overdorf, R., Troncoso, C., Greenstadt, R., & McCoy, D. (2018). Under the underground: Predicting private interactions in underground forums. *arXiv preprint*, arXiv:1805.04494.

Pastrana, S., Thomas, D. R., Hutchings, A., & Clayton R. (2018). CrimeBB: Enabling cybercrime research on underground forums at scale. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW '18*, 1845 – 1854.

Pastrana, S., Hutchings, A., Thomas, D., & Tapiador, J. (2019). Measuring e-whoring. In *Proceedings of the Internet Measurement Conference*, 463 – 477.

Pete, I., Hughes, J., Chua, Y. T., & Bada, M. (2020). A social network analysis and comparison of six dark web forums. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 484 – 493.

Potter, G. R. (2017). Real gates to virtual fields: Integrating online and offline ethnography in studying cannabis cultivation and reflections on the applicability of this approach in criminological ethnography more generally. *Methodological Innovations*, 10(1), 1 – 11.

Rodham, K., Gavin, J., Lewis, S., Bandalli, P., & St. Denis, J. (2016). The NSSI paradox: Discussing and displaying NSSI in an online environment. *Deviant Behavior*, 37(10), 1110 – 1117.

Siu, G. A., Collier, B., & Hutchings, A. (2021). Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum. In *2021 IEEE European Symposium on Security and Privacy Workshops*, 191 – 201.

Smirnova, O., & Holt, T. J. (2017). Examining the geographic distribution of victim nations in stolen data markets. *American Behavioral Scientist*, 61(11), 1403 – 1426.

Yip, M., Webber, C., & Shadbolt, N. (2013). Trust among cybercriminals? Carding forums, uncertainty and implications for policing. *Policing and Society*, 23(4), 516 – 539.

Naomi Lintvedt and Kristin Bergtora Sandvik

48 Robots

Abstract: The use of smart robots raises questions about their societal roles, human interaction, regulation, and impact on our understanding of crime and control. This chapter introduces three key issues related to digital criminology: bad robots, anthropomorphic robots, and crimes against robots. First, it addresses the use of robots in law enforcement, highlighting concerns about excessive or lethal force. Second, it explores the balance between caregiving roles and unethical surveillance by anthropomorphic robots. Third, it addresses violence against robots and the potential need for their protection. These issues necessitate reevaluating the frameworks governing the use and treatment of robots.

Keywords: android, anthropomorphism, humanoid, human–robot interaction, killer robot

Introduction

This entry explores so-called embodied smart robots, focusing on how the interface of software and hardware produces robots as physical objects acting upon the world with some degree of autonomy. The digital transformation of crime, the evolving capabilities of artificial intelligence, contentions around automation (see Automation by Mann), experimentation and prediction, and surveillance practices are central topics in digital criminology. The embodied robot is where all these tools, innovations, and practices meet and act upon the world. So far, robots have been the subject of limited critical attention from criminologists. To that end, the chapter provides a contextual description of robot attributes and capabilities before articulating a set of problematizations of interest for future criminological research.

What is a robot?

The standardization organization Institute of Electrical and Electronics Engineers (IEEE, 2023) classifies 18 different types of robots: Aerospace robots (e. g., the Mars rovers), aquatic robots, autonomous vehicles, consumer robots (e. g., robotic vacuum cleaners and pet robots), delivery robots, disaster response robots, drones, educational robots (e. g., robot kits to assemble and program), entertainment robots, exoskeletons (i. e. wearable robotic suits that help move the user's body), humanoid robots, industrial robots, medical robots (e. g., surgical robots), military and security robots, research robots, service robots, social robots (e. g., Paro a robotic seal), and telepresence robots (e. g., robots enabling remote presence). The classifications are overlapping; for example, a consumer robot may also be a social robot or a humanoid robot.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-049

Although robots are the topic of research and regulatory debates, there is no comprehensive legal definition of robots. Neither is there a common definition of robots in technical standards. The International Organization for Standardization (ISO, 2021, para 3.1) defines a robot as a "programmed actuated mechanism with a degree of autonomy to perform locomotion, manipulation or positioning." Hence, robots require a degree of autonomy, defined as "the ability to perform intended tasks based on current state and sensing, without human intervention" (ISO, 2021, para 3.2). The description offered by *The Encyclopaedia Britannica* (2023) sums up the vagueness of what a robot is: "any automatically operated machine that replaces human effort, though it may not resemble human beings in appearance or perform functions in a humanlike manner." In short: robots come in many different shapes, and many do not look like robots from popular culture—and we may fail to recognize them as robots.

How do robots interact with the world?

To be able to interact and adapt to their environment and human behavior, smart robots need to observe their surroundings via various sensors such as visual cameras, thermal cameras, audio, radar, Lidar (light detection and ranging), Wi-Fi detection, and more. To enhance human–robot interaction (HRI), robots are often fitted with facial-, voice-, and emotion recognition technologies. Robots collect massive amounts of data which AI further processes for them to learn and adapt. For example, the consumer robot dog 'Aibo' will adapt its behavior to its user based on the information it gathers through interaction and will develop a 'personality' based on these data. Thus, robots have enormous surveillance capacities which combined with their presence in our daily life give them access to both public and private spaces. Importantly, as noted by Calo and colleagues, "robotics combines, for the first time, the promiscuity of data with the capacity to do physical harm" (Calo et al., 2016). The combination of their physical presence and complex data processing capabilities raises intriguing questions about what role robots should have in society, how they can and should interact with humans, how they should be regulated, and how they shape our understanding of crime and crime control. Caveating the issue of drones, this chapter offers a brief inventory of these issues and sets out an agenda for further research.

Robots as socio-technical imaginaries: (popular) culture

From a perspective of digital criminology, fundamental questions pertain to the usages of robots for criminal actions; the thresholds for when design problems and malfunction become the source of criminal responsibility; and, in the future, whether sentient robots may one day have the capacity to hold criminal responsibility. Underpinning all

these questions, while embodied robots are tangible, physical objects, they are also produced through problem framings and sociotechnical imaginaries. Sociotechnical imaginaries are collectively held, institutionally stabilized, and publicly performed visions of how technology can produce a desirable future (Jasanoff, 2015). These visions are premised on the notion that technology can produce truth and bring order and progressive social change. This way, socio-technical imaginaries are also "performances of power" (Shelby, 2021).

Where do robots come from?

The understanding of what a robot is, what it looks like, and what it can do is defined by mostly popular culture. The etymology of the word 'robot' is from the Czech word 'robota,' meaning servitude or forced labor. It first appeared in Karel Capek's play *R.U.R., Rosumovi Univerzální Roboti* (Rossum's Universal Robots) from 1921. Yet, the concept of robots began much before with *automata* in ancient Greek, many made as mechanical humans or animals. Today, the word 'robot' suggests hostile robots from movies such as *Terminator* and *RoboCop*, but also friendly ones like R2D2 and Wall-E. In popular culture, robots are often cast as a menace to humans, or gaining human-like consciousness such as in *Blade Runner* and *Westworld* thereby raising the question of what it means to be human. However, since the imaginaries of robots are influenced by popular culture, this also differs across cultures. Perceptions of and approaches to robots vary cross-culturally with significant variations in acceptance. In turn, this acceptance shapes the degree to which robots are seen as solutions to political problems. For example, in Japan with its advanced robotics industry, robots are often featured in animations (manga and anime) as part of the household and co-existing with humans (e.g., the robot boy Tetzuwan Atom (Astro Boy) and the robot cat Doraemon which travels back in time to be a companion to a boy and stays with his family), or forming a bio-mechanical symbiosis with humans (e.g., the fighting robots Mobile Suit Gundam and Neon Evangelion which are piloted by teenagers to protect civilization). Robots are cast as the solution to the Japanese demographic crisis with an aging and decreasing Japanese population, supporting the traditional household, instead of more human resources by immigration (Robertson, 2014).

What does it mean to regulate robots? From bans to liability

Attempts to imagine and control moral transgressions by robots are central to the socio-technical imaginary. For digital criminology, this imaginary is central to evolving ideas about robots as perpetrators, victims, and vehicles for crimes. Science fiction au-

thor Asimov sought to create an ethical system for human–robot interaction and formulated three laws of robotics in the short story 'Runaround' (1942):

- A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- A robot must obey orders given to it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.
- Asimov later added the zeroth law: A robot may not harm humanity, or, by inaction, allow humanity to come to harm.

His laws have had tremendous normative and policy influence on the robotics field, even if the purpose of many robots is to harm people (e. g., military robots). It is important to understand that the regulation of robots goes beyond the law to include design and architecture, market forces, and social norms. Robot regulation poses the challenge of keeping up with technological advances; striking a balance between stimulating innovation and the protection of fundamental rights and values; deciding whether to affirm prevalent social norms or nudge social norms in a different direction; and finally, how to balance effectiveness versus legitimacy in techno-regulation (Leenes et al., 2017).

With respect to the criminalization of specific types of robots, the perhaps globally most well-known robot regulation initiative is the so-called Stop Killer Robots campaign, which started in 2012. Citing the need to fight against digital dehumanization and ensure human control in the use of force, the campaign calls for a ban on lethal autonomous weapons (https://www.stopkillerrobots.org/). The most widespread criminalization of robots concerns child sex robots, in the form of robots embodying realistic and anatomically correct representations of children produced for sexual exploitation. In many jurisdictions, such robots are now banned (Darling, 2021). Some countries, such as Norway and the UK have decided that childlike sex dolls can be classified as child pornographic material and thus fall under existing criminal law. In contrast, others, like Australia, Germany, and Denmark have adopted new laws criminalizing childlike sex dolls (Loibl et al., 2023).

Nevertheless, today, most regulations of robotics are safety regulations for industrial robots that are used for specific purposes in enclosed areas, often with no human interaction, for example, robots used in car manufacturing plants. There is no specific regulation of other robots or human–robot interaction. The regulation is fragmented and requires applying a variety of rules on privacy, data protection (see Privacy and Data Protection by Bygrave), AI, product safety, consumer protection, liability, etc. Some robots fall under the requirements of, e. g., electronic products and must comply with safety standards. However, these regulations do not address other adverse effects that may occur in human–robot interaction and do not address concerns about the use of, e. g., facial recognition and emotion recognition technologies or how robots affect

the physical and psychological well-being of persons who are using or interacting with robots.

A central aspect of the imagination of robots concerns the role of robots: some see robots as tools or slaves, while others point to the dystopia of robots as overlords of humans. A common concern, as reflected in Asimov's laws, is the potential violence from robots. Embodied robots are heavy and mechanical and have the potential to physically harm humans. This can be accidental because of malfunction, but also caused by hacking where third parties overtake the functionality of the robots. There are numerous incidents with autonomous vehicles (which may be labeled as a type of robot) that have been hacked or that cause accidents due to their AI system's inability to detect and predict pedestrian behavior. At the same time, the existence of deeply problematic artifacts such as child-sex robots complicates the relationship between robots and harm and how we understand violence against robots. In the next section, we delineate three issues in need of further attention.

Emergent issues: bad robots, anthropomorphic robots, and crimes against robots

The first issue concerns *bad* robots. Function creeps create new questions with respect to what constitutes legal and illegal robot use. Some robots can be classified within several use categories. Robots from one field also spill into other fields where they are used to deploy violence. This includes technological transfers from industry to the police and from the military to the police. For example, robots can be useful to inspect suspicious packages or to diffuse explosives. However, the marketing of robots may intentionally go beyond their intended purpose. A robot developed for inspection purposes in industrial environments—'Spot' (see Ethics by Markham)—is also being marketed as a tool for public safety and to "de-escalate hostage scenarios" (Boston Dynamics, 2023a). Police forces in some US states have already started using Spot in, for example, house searches, prompting criticism from civil rights organizations and raising ethical concerns (Yunus and Doore, 2021). Since the robot has limited functionality, the purpose of use seems to be to intimidate and scare people rather than the robot appeasing the situation. Other police forces are discussing using military-purpose robots or equipping robots with lethal force. Some of the leading manufacturers of general-purpose robots have called out against efforts to weaponize commercially available robots (Boston Dynamics, 2023b). The case of so-called 'killer robots' in warfare is already problematic, and extending their use to law enforcement will raise further issues on autonomy, human control, and the use of force (Sandvik et al., 2014). This includes how law enforcement robots should be permitted to react to hostile situations, including the use of lethal force. Law enforcement is allowed to use force to protect life, including threats to their own life. A robot can be permanently damaged, but it is not alive. If a robot is allowed to react, based on data and AI, in the same way as human police of-

ficers its use of force may be excessive. Thus, the deployment of robots in law enforcement may warrant other rules of action than for the human police force.

A second type of issue concerns the *balance* between using robots to offer care and when it veers over into illegal and unethical surveillance, and what human likeness means for this balance. Robots are given work that is dull, dirty, and dangerous. Robots may be designed as mechanical shapes, i. e., delivery robots that are boxes with wheels, but many robots have anthropomorphic or zoomorphic designs such as androids or pets or cute designs. When robots are developed to be companions to humans—for example, as a substitute for pets—this raises questions about human autonomy, personhood, privacy, and agency (see Agency by Krasmann). The ethical concern is that anthropomorphic robots may lead to deception of the user and emotional dependency on a mechanical being. From a privacy perspective, a friendly robot can lead the user to disclose more personal information and private behavior than mechanical robots. However, research also shows how people bond with mechanical robots such as vacuum cleaners, e. g., naming and personalizing the robots, and military robots, e. g., soldiers performing memorial services for robots (Carpenter, 2013). Replicating human posture and gait in robotics is difficult, and most robots are made without legs as these tend to topple over due to their weight and lack of balance. However, robots are often portrayed as having human-like features, not only in science fiction but in research and industrial projects. For example, Tesla has presented a faceless robot with a human-like body and Boston Dynamics promotes their robots with videos of robots doing acrobatics and parkour. The latter is a promotional clip with careful scenography and editing, exaggerating the current abilities of the robots, and propagating a sociotechnical imaginary of robots as a type of super-human (Moses and Ford, 2021).

The third point concerns crimes *against* robots. There is not only concern about violent robots, but violence against robots. When robots are 'in the wild,' several experiments show that people, especially children, act aggressively against them. Research suggests that children engaging in the abuse of robots did so because they were curious about the robot's reactions or enjoyed abusing it and considered it human-like. Although a robot may be human-like or pet-like, violence against a robot will be like damaging a property: you may be held liable if you damage someone else's property, but you are free to damage your own property. If a robot should be treated differently from property it would be based on the notion that a robot is a sentient being or that it has its own rights. Various strategies for protecting robots beyond the property justification have been proposed (Mamak, 2023): a highly contested proposition involves giving robots legal personhood and moral rights (Coeckelbergh, 2010). A different approach involves criminalizing public violence against robots as a violation of public morality (Mamak, 2022). As noted above, the most difficult current issue concerns child sex robots, involving multilayered forms of violence. The existence of such robots is considered morally repugnant and exploitative. The abuse of such robots and the possible impact on real children adds further dimensions.

Conclusion

This chapter has offered a contextual description of robots and the challenges of regulating robots. Focus has been given to the global diversity of robots as sociotechnical imaginaries in a rapidly moving regulatory field, and to identify some issues of specific relevance for digital criminology. The chapter has focused on the use of robots for violence and control, issues of surveillance and autonomy arising from the design and deployment of robots as care objects, but also the more generalized anthropomorphic ascription of mechanical robots and finally the emergent class of robots as victims of crimes and violence.

Pointers for further research

To contribute to the field of digital criminology—or perhaps even a subfield of 'robot criminology'—more qualitative use cases are needed across the array of areas discussed in this chapter, including developments in 'robot crimes,' criminal law and policing practice, and with a view to analyze R & D processes and uptake of new robots in law enforcement. Important questions also arise around our future with robots, how human–robot interactions may engender new crimes, and one day, whether fully, autonomous sentient robots will acquire the capacity to be held criminally responsible for their actions and inactions.

Suggested reading

Calo, R., Froomkin, A. M., & Kerr, I. (eds.) (2016). *Robot Law.* Cheltenham: Elgar.

Darling, K. (2021). *The New Breed: What Our History with Animals Reveals about Our Future with Robots.* New York: Henry Holt and Company.

References

Asimov, I. (1950). *I, Robot.* New York: Gnome Press.

Boston Dynamics. (2023a). *Spot.* Government & Public Safety Kit. https://bostondynamics.com/wp-content/uploads/2023/10/spot-government-public-safety-kit.pdf

Boston Dynamics. (2023b). *General purpose robots should not be weaponized.* https://bostondynamics.com/news/general-purpose-robots-should-not-be-weaponized/

Calo, R., Froomkin, A. M., & Kerr, I. (eds.) (2016). *Robot Law.* Cheltenham: Elgar.

Carpenter, J. (2013). *The Quiet Professional: An investigation of US military Explosive Ordnance Disposal personnel interactions with everyday field robots.* PhD dissertation, University of Washington.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209 – 221. doi:10.1007/s10676 – 010 – 9235 – 5

Darling, K. (2021). *The New Breed: What Our History with Animals Reveals about Our Future with Robots.* New York: Henry Holt and Company.

Encyclopaedia Britannica. (2023). Robot. https://www.britannica.com/technology/robot-technology.

Institute of Electrical and Electronics Engineers [IEEE]. (2023). *Types of Robots. Categories frequently used to classify robots.* https://robotsguide.com/learn/types-of-robots

International Organization for Standardization [ISO]. (2021). ISO 8373:2021 Robotics – Vocabulary

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasanoff & S.-H. Kim (eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (pp. 1 – 33). Chicago: University of Chicago Press.

Leenes, R., Palmerini, E., Koops, B.-J., Bertolini, A., Salvini, P., & Lucivero, F. (2017). Regulatory challenges of robotics: Some guidelines for addressing legal and ethical issues. *Law, Innovation and Technology*, 9(1), 1 – 44. https://doi.org/10.1080/17579961.2017.1304921

Loibl, E., van der Aa, S., Hendriks-Lundh, M., & Niemark, R. (2023). Exploring different national approaches to prohibiting childlike sex dolls. *Maastricht Journal of European and Comparative Law*, 30(1), 63 – 82. 1023263X231176908.

Mamak, K. (2022). Should violence against robots be banned? *International Journal of Social Robotics*, 14(4), 1057 – 1066.

Mamak, K. (2023). *Robotics, AI and Criminal Law: Crimes against Robots.* London: Routledge.

Moses, J., & Ford, G. (2021). See Spot save lives: Fear, humanitarianism, and war in the development of robot quadrupeds. *Digital War*, 2, 64 – 76. https://doi.org/10.1057/s42984-021-00037-y

Robertson, J. (2014). Human rights vs. robot rights: Forecasts from Japan. *Critical Asian Studies*, 46(4), 571 – 598. https://doi.org/10.1080/14672715.2014.960707

Sandvik, K. B., Marsh, N., & Mirshahi, M. (2014). The struggle to ban killer robots. *Bulletin of Atomic Scientists*, 7 May. https://thebulletin.org/2014/05/the-struggle-to-ban-killer-robots/

Shelby, R. (2023). Technology, sexual violence, and power-evasive politics: Mapping the anti-violence sociotechnical imaginary. *Science, Technology, & Human Values*, 48(3), 552 – 581. https://doi.org/10.1177/01622439211046047

Yunus, A., & Doore, S. A. (2021). Responsible use of agile robots in public spaces. *IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*, 1 – 5. https://doi.org/10.1109/ETHICS53270.2021.9632682

Pamela Ugwudike

49 Sentencing and risk assessment algorithms

Abstract: Digital technologies in the form of Artificial Intelligence (AI) models, particularly machine learning algorithms, are increasingly influencing decision-making in justice systems around the world. This contribution focuses on their role in partially automating judicial decision-making. It is structured around three sections. First, it outlines the origins of the technologies and describes how they influence sentencing. Second, it considers the justifications offered by their proponents, and third, it critically explores their implications for justice, linking the analysis to discourses about their capacity to reproduce biases and their lack of transparency. Finally, the chapter provides recommendations on how best to remediate pressing challenges. Although it focuses on developments in the UK, it refers to other jurisdictions where relevant.

Keywords: Artificial Intelligence, risk assessment algorithms, predictive algorithms

Introduction

Although its definition is varied and contested, AI is commonly depicted as "general purpose digital technologies that enable machines to do highly complex tasks effectively" (Hall and Pesenti, 2017; see also Matheis and Kingdon, 2023). This definition encompasses several techniques and technologies including machine learning algorithms that predict risk of recidivism and can influence how judges evaluate risk to determine the appropriate sentence. Some of the risk assessment algorithms are basic logistic regression models which became widespread across the UK and other countries from the 1990s onwards, whilst several now rely on more advanced machine learning techniques.[1]

According to the Sentencing Council (2023) which provides guidelines on sentencing in England and Wales, "A sentence is the punishment a judge or magistrate [the sentencer] decides should be given to someone who has been convicted of a crime." Risk assessment algorithms (see Artificial Intelligence by Van Brakel) are trained to analyze data and detect patterns which inform decision-making in the context of, for example, crime risk prediction and sentencing (see Prediction by Ķīlis, Gundhus, and Galis). As advanced computational technologies, their decision-making is, so the argument goes, more objective than that of human actors, and their risk predictions can be considered during sentencing. This chapter's contribution lies in its analysis of the origins of such sentencing technologies, how they influence sentencing practices, and the

1 See Berk and Bleich (2013) for a detailed description of statistical approaches to predicting risk of recidivism.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-050

justifications offered by their proponents. It also provides insights on the implications of the technologies for justice and connects the analysis to discourses about their capacity to reproduce bias and their lack of transparency (see Bias by Oswald and Paul). The chapter concludes with recommendations on how best to remediate pressing challenges.

AI and sentencing

Known colloquially as risk assessment tools, risk assessment algorithms (some of which are machine learning models and are as such increasingly described as AI) are deployed in justice systems. The technologies are used at various phases of the criminal justice process including the post-conviction, pre-sentence stage when judges draw on various sources of information about a defendant to craft a sentence. Several risk assessment algorithms form part of the criminal justice AI infrastructure. They include sections of the Offender Assessment System—OASys (UK), the Correctional Offender Management Profiling for Alternative Sanctions—COMPAS (USA), the Level of Service Inventory-Revised (LSI-R) and the Level of Service/Case Management Inventory—LS/CMI (Australia, Canada, Sweden, Switzerland, and UK).

A brief history

There are similarities between the risk assessment algorithms currently playing a key role in sentencing across several Western jurisdictions and the statistical approach established by Burgess (1928) for parole decision-making. Burgess analyzed data from criminal justice populations to identify factors which he believed were linked to parole outcomes. Then, he tested the factors on a retrospective sample and used them to design a risk assessment tool for determining risks of parole violation (see also Kemshall, 2019).

In an extension of this actuarial approach to risk prediction, some of the currently used risk assessment algorithms deploy machine learning techniques including statistical methods such as logistic regression models (Kemshall, 2019; Raynor, 2019). Advanced machine learning techniques for analyzing more risk factor variables than regression models, are also used to predict crime risks (see Howard, 2017; Law Society, 2019).

Generally, the process of designing risk assessment algorithms involves training the technologies (such as a logistic regression model), using historical data, so they learn how to, (1) process data from defendants and other risk subjects, (2) detect patterns e.g., risk factors (which are sometimes pre-defined), and (3) predict risk (Moore, 2015). Such risk factors are variables said to correlate with recidivism and they include, for example, demographic information (typically, age, gender, and criminal history); socioeconomic and lifestyle factors (accommodation, educational and employment sta-

tus, family relationships, leisure pursuits, alcohol, and drug misuse); attributed personality, cognitive and behavioral characteristics (behavior, thinking, attitudes) (Ministry of Justice, 2019). Individuals whose profile indicate the presence of these factors, particularly those weighted as strongly correlated with reconviction (typically within two years of completing a custodial or community sentence), will attract higher risk scores than others. The predetermined risk factors are derived from various theories. Examples are criminological and sociopsychological perspectives that depict factors such as familial and social-environmental circumstances as well as rational choice as the *causes* of recidivism (see generally, Bonta and Andrews, 2017). It is nevertheless argued that they are in fact only possible *correlates* (Prins and Reich, 2018).

Automating sentencing practice

The ways in which the risk scores produced by such digital predictive technologies influence sentencing varies across jurisdictions. In Western countries such as the UK, US, and Canada, the scores are embedded in the reports that sentencers consult to determine the appropriate sentence. In the UK, a Pre-Sentence Report (PSR) as defined by Section 158 (1) of the Criminal Justice Act 2003 and Section 31 (1) of the Sentencing Act, is written, "with a view to assisting the court in determining the most suitable method of dealing with an offender." Further, according to the UK's Sentencing Council (2023a), a fundamental role of the probation service in court is "preparing pre-sentence reports for courts, to help them select the most appropriate sentence." The reports allow the courts to meet their legal obligation to take various risks into account during sentencing.

All PSRs should outline both the risk scores ascribed to the convicted person and the recommendations of the report writer (the probation officer) regarding a sentence that is commensurate with the predicted risk. Judges are expected to consider all that information when crafting the appropriate sentence. Although further research is required, official statistics point to a high concordance between PSR recommendations and sentencing decisions: "89 % of immediate custodial sentences proposed in PSRs resulted in that sentence being given in the year ending December 2022" (Ministry of Justice, 2023).

In general, through PSRs, risk assessment algorithms provide information with which the courts can impose sentences that restrict civil rights and liberties, occasionally beyond set tariffs. In the US they can influence sentencing in capital cases (Kehl et al., 2017). Yet, the scores and classifications produced by risk assessment algorithms can never state with absolute certainty the likely or exact risk posed by an individual. A risk score only ever indicates that some people fitting a certain profile may reoffend and cannot categorically conclude that the defendant is one of those who will (see Hamilton and Ugwudike, 2023). A reason for this is that risk assessment algorithms predict individual risk on the basis of risk factors derived from statistical analysis of data from various sources including criminal justice populations (see, for example, Howard

and Dixon, 2012). This lack of individualization heightens the probability of unfair profiling. In short, the algorithms generate predictions about group risks not individual risk. Yet, they are used for the latter in high stakes sentencing processes.

Implications of AI-driven sentencing: claims and counterclaims

The high level of mathematical quantification underpinning predictive algorithms such as OASys (UK) and COMPAS (US), create a veneer of scientific objectivity that ostensibly transcends professional judgment which has long been depicted as fundamentally prone to biases and prejudices (Grove and Meehl, 1996). Indeed, proponents and state procurers of predictive algorithms typically emphasize their value and prospects, using optimistic frames which depict them as capable of analyzing large-scale data objectively whilst reducing human bias, optimizing cost-effective resource allocation, and enhancing institutional efficiency (see Huq, 2019; Lavorgna and Ugwudike, 2021). Nevertheless, predictive algorithms pose challenges for justice. This chapter now explores the problems of bias and lack of transparency, both of which undermine key antecedents of justice, particularly non-discriminatory practice and procedural due process.

Racial bias

While there are several potential conduits of bias, a main source is the type of data on which the algorithms rely for prediction. An example is administrative data on criminal justice populations, such as arrest and conviction records compiled by police and court services (Moore, 2015). Such data can contain records of racially biased decision-making. Considering the case of arrest records, certain minorities, particularly Black people, are vulnerable to racially biased arrests in Western jurisdictions such as the UK (Shiner et al., 2018) and the US (e.g., Richardson et al., 2019) where risk assessment algorithms already influence sentencing decisions. As the technologies are designed to interpret historical arrest data as proxies for crime and objective risk predictors, they are likely to infer from the higher arrest rates of affected minorities that this group is more prone to criminality than others. This can prompt the tools to inflate their risk scores, reproducing biases embedded in the historical data (Hao and Stray, 2019; Mayson, 2019; Ugwudike, 2020). This also highlights the problem of lack of individualization involving predictions made on the basis of group risks rather than individual risk.

Risk predictions are only ever as good as the data on which the technologies rely: where the data comprises records of discriminatory decision-making, the predictions will reproduce the bias. If the predictions inform sentencing decision-making, principles of justice such as fairness and impartiality will be undermined.

Cultural bias

A related issue is that AI applications including risk assessment algorithms are not created in a normative or cultural vacuum. They reflect the westernized norms, values, and ideologies of their designers and procurers and can lack cultural relevance amongst non-Western communities. Indigenous groups in Australia and Canada where sentencing technologies are being deployed represent examples. In the case of Ewert v Canada Correctional Service (2018), for example, Jefferey Ewert, an indigenous Canadian was denied parole on the basis of high-risk predictions generated by risk assessment algorithms. On appeal, the Supreme Court of Canada agreed with Ewert that the technologies did not consider cultural factors relevant to indigenous communities and could instead label the factors as risky, exposing affected defendants to risk inflation. During sentencing, such risk amplification would undermine the principles of impartiality and fair treatment vital for the equitable administration of justice.

Socioeconomic bias

Bias based on adverse socioeconomic circumstances represents yet another challenge associated with the risk assessment algorithms that influence sentencing. As already noted, some of the commonly used risk assessment algorithms are designed to generate predictions based on predetermined risk factors selected by their designers. These include signs or indices of socioeconomic deprivation. Unsuitable or unstable accommodation, unemployment, poor educational attainment, and lack of engagement in lawful leisure activities are examples (Ministry of Justice, 2019). However, these can operate as proxies for deprivation, rendering socioeconomically disadvantaged defendants vulnerable to higher risk scores than they deserve (Van Eijk, 2016), which, again, undermines principles of justice.

Gender bias

Gender-related bias, which has also been associated with risk assessment algorithms, is another factor that poses implications for principles of justice, particularly procedural due process. Studies and commentators note that some of the tools have been trained and validated using data from young male populations, calling into question their suitability for women (e. g., van der Knaap et al., 2012). Concurrent work has also found that risk factors such as socioeconomic circumstances expose women to risk inflation as they tend to be more affected by inequalities in those areas compared with men and would as such appear riskier (Hamilton, 2019).

In sum, predetermined risk factors such as criminal history and socioeconomic status which inform algorithmic risk scores, can operate as proxies for race, gender,

and socioeconomic status. In doing so, they introduce proscribed extrajudicial characteristics into sentencing decisions and undermine both procedural due process (see Article 6 of the UK's Human Rights Act 1998) and the right to non-discrimination (Article 14).

Transparency deficits

Lack of transparency represents yet another limitation associated with AI technologies in general (Oswald et al., 2022) including the risk assessment algorithms that inform sentencing decisions (Kehl et al., 2017). Given the level of advanced computational analysis involved, algorithms can evolve into black boxes. They can become esoteric or instructible even to their designers, who are in any event, protected by trade secret laws, meaning that they do not have to reveal the contents of their code. Even if sentencing technologies were transparent and open-source models, financial and technical resources for meaningful algorithm audit are still required. Such lack of transparency obfuscates conduits of bias and limits accountability. It also denies defendants the opportunity to rebut ascribed algorithmic risk scores, with adverse implications for the due process principle that is central to the fair administration of justice (see Article 6 of the Human Rights Act 1998; Ugwudike, 2020).

Relevance to criminology

The challenges presented by risk assessment algorithms used for sentencing fall within the purview of criminology, including the field of digital criminology, and the strand of the discipline that is concerned with social harms. Whilst challenges such as bias and lack of transparency are not proscribed by law, they produce serious social harms. Consider for instance, the potential implications of a high risk of recidivism score informed by racial, gender, or socioeconomic bias. It heightens the risk of an unwarranted custodial sentence which punishes the affected defendant for a future crime, denying them their civil rights and liberties. It also poses adverse consequences for their family (Ugwudike and Fleming, 2023). Post-release, their conviction undermines liberal principles of distributive justice; it unfairly restricts access to key resources, from accommodation and education to employment (see generally Corda et al., 2023).

Another area of criminological relevance is the capacity of risk assessment algorithms to undermine judicial autonomy. This can occur in two ways. One is via the expectation that judges should rely on decision-making tools commissioned by policymakers, blurring the constitutional boundary between the executive and judicial arms of government. Another is through the phenomenon of algorithm appreciation (Logg et al., 2019) or 'automation bias' (Goddard et al., 2012): algorithmic technologies extend the trend in sentencing that involves privileging data-driven, non-personalized, knowl-

edge about defendants and imposing sentences based on similarities with a group rather than individual actions.

Conclusion and AI futures

The challenges of AI in sentencing compel us to reflect on how best to counteract them and design responsible models for the future. While some call for outright abolition (e.g., Schwerzmann, 2021), others pragmatically propose robust legal frameworks (e.g., Martini 2020). In the UK, official efforts to mitigate the challenges in the public sector include the introduction of an Algorithmic Transparency Standard which is an auditing system that has been trialled with the police (Oswald et al., 2022). Apart from audits, other technical remedies such as improving explainability (Zeng et al., 2015) and debiasing data have been proposed and trialled (Skeem and Lowenkamp, 2020). Some reject the very notion of debiased or unbiased data and argue quite persuasively that data always contains and reflects the choices, assumptions, and interpretations made during collection and processing (see Gitelman and Jackson, 2013; Ugwudike 2022). Remedial strategies must therefore focus on understanding the type of bias embedded in data and how best to mitigate them (Kaufmann, 2023).

Beyond technical mitigations, proffered structural remedies include equitable distribution of digital capital (van Dijk, 2005) in the form of access to technologies and design skills. At present, much of the digital capital required for constructing and auditing AI technologies resides in a few, centralizing power amongst them whilst excluding others who currently bear the culminative ethical burden of the technologies (Petrozzino, 2021; Ugwudike, 2022).

Main takeaways

- Risk assessment tools in the form of machine learning algorithms can directly influence sentencing decisions.
- Justifications offered for their use focus on assumptions of scientific objectivity, bias elimination, cost effectiveness, institutional efficiency, and public protection.
- Bias and lack of transparency are key challenges affecting the algorithms and they pose implications for justice.
- Proffered remedial strategies include legal and regulatory frameworks.
- Third-party audits have also been proposed to anticipate unintended consequences.
- Structural level remedies include access to digital capital to reverse the uneven distribution of risks and benefits.

Suggested reading

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals and it's biased against Blacks.* ProPublica, 23 May. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Hamilton, M., & Ugwudike, P. (2023). A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up. https://theconversation.com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594

Hao, K., & Stray, J. (2019). Can you make AI fairer than a judge? Play our courtroom algorithm game. *MIT Technology Review.* Available at: https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm

Kehl, D., Guo, P., & Kessler S. (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing.* Responsive Communities Initiative, Berkman Klein Centre for Internet & Society: Harvard Law School.

Law Society. (2019). *Algorithm Use in the Criminal Justice System Report.* Available at: https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report

Ugwudike, P. (2020). Digital prediction technologies in the justice system: The implications of a 'race-neutral' agenda. *Theoretical Criminology*, 24(3), 482 – 501.

References

Angwin J., & Larson, J. (2016) Bias in criminal risk scores Is mathematically inevitable, researchers say. https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

Association for Computing Machinery (ACM), US Public Policy Council, and ACM Europe Council 2017

Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behaviour: A comparative assessment. *Criminology & Public Policy*, 12, 513 – 544.

Bonta, J., & Andrews, D. A. (2017). *The Psychology of Criminal Conduct* (6th ed). New York, NY: Routledge.

Burgess, E. (1928). Factors making for success of failure on parole. *Journal of Criminal Law and Criminology*, 19, 239 – 306.

Corda, A., Rovira, M., & Zand-Kurtovic, E. van 't. (2023). Collateral consequences of criminal records from a cross-national perspective: An introduction. *Criminology & Criminal Justice*, 23(4), 519 – 527.

Ewert v. Canada, 2018 SCC 30, [2018] 2 S.C.R. 165. https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/17133/index.do

Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (ed.), *"Raw Data" is an Oxymoron* (pp. 1 – 14). Cambridge, MA: MIT Press.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of American Medical Informatics Association*, 19(1), 121 – 127.

Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293 – 323.

Hall, W., & Pesenti, J. (2017). *Growing the Artificial Intelligence Industry in the UK.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

Hamilton, M. (2019). The sexist algorithm. *Behavioral Sciences and the Law*, 37(2), 145 – 157.

Hamilton, M., & Ugwudike, P. (2023). A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up. https://theconversation.com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594

Hao, K., & Stray, J. (2019). Can you make AI fairer than a judge? Play our courtroom algorithm game. *MIT Technology Review.* Available at: https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm

Howard, P. (2017). Offender group reconviction scale. In J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, & Z. Hamilton (eds.), *Handbook of Recidivism Risk/Needs Assessment Tools* (pp. 231 – 242). Hoboken, NJ: John Wiley & Sons, Ltd.

Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASys Violence Predictor: Advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior*, 39(3), 287 – 307.

Huq, A. Z. (2019). Racial equity in algorithmic criminal justice. *Duke Law Journal*, 68, 1043 – 1134.

Kaufmann, M. (2023) *Making Information Matter. Understanding Surveillance and Making a Difference.* Bristol: Bristol University Press.

Kemshall, H. (2019). A critical review of risk assessment policy and practice since the 1990s. In P. Ugwudike, H. Graham, F. McNeill, P. Raynor, F. Taxman, & C. Trotter (eds.), *Routledge Companion to Rehabilitative Work in Criminal Justice* (pp. 220 – 231). Abingdon: Routledge.

Kehl, D., Guo, P., & Kessler, S. (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing.* Responsive Communities Initiative, Berkman Klein Centre for Internet & Society: Harvard Law School.

Law Society. (2019). *Algorithm Use in the Criminal Justice System Report.* Available at: https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report

Lavorgna, A., & Ugwudike, P. (2021). The datafication revolution in criminal justice: An empirical exploration of frames portraying data-driven technologies for crime prevention and control. *Big Data & Society*, 8(2).

Logg, J., Minson, J., & Moore, D. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behaviour and Human Decision Processes*, 151, 90 – 103.

Martini, M. (2020). Regulating algorithms: How to demystify the alchemy of code? In M. Ebers & S. Navas (eds.), *Algorithms and Law* (pp. 100 – 135). Cambridge: Cambridge University Press.

Mattheis, A. A., & Kingdon, A. (2023) Moderating manipulation: Demystifying extremist tactics for gaming the (regulatory) system. *Policy & Internet*, 15(4), 478 – 497.

Mayson, S. G. (2019). Bias in, bias out. *The Yale Law Journal*, 128(8), 2122 – 2473.

Ministry of Justice. (2019). *Identified needs of offenders in custody and the community from the Offender Assessment System, 30 June 2018 Ad Hoc Statistics.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/815078/oasys-needs-adhoc-stats.pdf

Ministry of Justice. (2023). *Offender Management Statistics Bulletin, England and Wales Quarterly: October to December 2022, and Annual: Calendar Year 2022.* https://assets.publishing.service.gov.uk/media/644915fcf126830011ca6882/OMSQ_Q4_2022.pdf

Moore, R. (2015). The Offender Assessment System (OASys) and the 2009 – 2013 research projects. In R. Moore (ed.), *Ministry of Justice Analytical Series: A compendium of research and analysis on the Offender Assessment System (OASys) 2009 – 2013.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf

Oswald, M., Chambers, L., Goodman, E. P., Ugwudike, P., & Zilka, M. (2022). *The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives.* Available at SSRN: https://ssrn.com/abstract=4155549 or http://dx.doi.org/10.2139/ssrn.4155549

Petrozzino, C. (2021). Who pays for ethical debt in AI? *AI Ethics*, 1, 205 – 208.

Prins, S. J., & Reich, A. (2018). Can we avoid reductionism in risk reduction? *Theoretical Criminology*, 22(2), 258 – 278.

Raynor, P. (2019). Risk and need assessment. Development, critics, and a realist approach. In P. Ugwudike, H. Graham, F. McNeill, P. Raynor, F. Taxman, & C. Trotter (eds.), *Routledge Companion to Rehabilitative Work in Criminal Justice* (pp. 210 – 219). Abingdon: Routledge.

Richardson, R., Schultz J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review*, 94(2), 192 – 233.

Schwerzmann, K. (2021). Abolish! Against the use of risk assessment algorithms at sentencing in the US criminal justice system. *Philosophy and Technology*, 34, 1883 – 1904.

Sentencing Council (2023) How sentencing works: What is a sentence? https://www.sentencingcouncil.org.uk/sentencing-and-the-council/how-sentencing-works/

Sentencing Council. (2023). Sentencing basics. https://www.sentencingcouncil.org.uk/sentencing-basics/#:~:text=The%20sentence%20imposed%20on%20an,case%20they%20are%20dealing%20with.

Sentencing Council. (2023a). Who does what in sentencing? (https://www.sentencingcouncil.org.uk/sentencing-and-the-council/who-does-what-in-sentencing/#Probation_Service

Shiner, M., Carre, Z., Delsol, R., & Eastwood, N. (2018). *The Colour of Injustice: 'Race', Drugs and Law Enforcement in England and Wales.* https://www.lse.ac.uk/united-states/Assets/Documents/The-Colour-of-Injustice.pdf

Skeem, J., & Lowenkamp, C. (2020). Using algorithms to address trade-offs inherent in predicting recidivism. *Behavioral Sciences & The Law*, 38(3), 259 – 278.

Ugwudike, P. (2020). Digital prediction technologies in the justice system: The implications of a 'race-neutral' agenda. *Theoretical Criminology*, 24(3), 482 – 501.

Ugwudike, P. (2022). Predictive algorithms in justice systems and the limits of tech-reformism. *International Journal for Crime, Justice, and Social Democracy*, 11(1), 85 – 99.

Ugwudike, P., & Fleming, J. (2023). Artificial Intelligence, digital capital, and epistemic domination on Twitter: A study of families affected by imprisonment. *Punishment and Society,* 25(1), 99 – 121.

van der Knaap, L. M., Alberda, D. L., Oosterveld, P., & Born M. P. (2012). The predictive validity of criminogenic needs for male and female offenders: Comparing the relative impact of needs in predicting recidivism. *Law and Human Behavior,* 36(5), 413 – 422.

van Dijk, J. A. G. M. (2005). *The Deepening Divide. Inequality in the Information Society.* London: Sage.

Van Eijk, G. (2016). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. *Punishment and Society*, 19, 463 – 481.

Zeng, J., Ustun, B., & Rudin, C. (2015). Interpretable classification models for recidivism prediction. *arXiv.* https://arxiv.org/abs/1503.07810

Helen M. Rand

50 Sex work

Abstract: This chapter does three things. It provides a definition of sex work, specifically outlining the meaning of digital sex work. Second, in order to understand digital sex work, the chapter unpacks the role of platforms within this labor market, and how sex workers interact with adult platforms. This highlights both the risks and opportunities for sex workers online, including the role of the internet in activism and advocacy. The final part of the chapter evaluates the regulation and legislation of digital sex work, drawing attention to the harms of further criminalization.

Keywords: sex work, digital sex workers, platforms, online regulation

The term sex work was coined by sex worker activist Carol Leigh in 1978, before the digitization of sex work, as part of a long history of activism and political organization to assert sex worker rights. For sex work activists and scholars, using the term sex work is part of a wider campaign to decriminalize sex work and reduce stigmatization. Recognizing selling sex and sex acts as *sex* and as *work*, frames sex work as a form of labor with rights and not a criminal act. It is a term that is used to incorporate a range of labor processes and work experiences, such as stripping, porn acting, telephone sex, webcamming, and direct, in-person sex work (Weitzer, 2010). Yet, it should be noted that not all people who exchange sex or sexual services define themselves as sex workers (see Blunt and Wolf, 2020 for a discussion on this).

The term sex work is now often used in academia, the media, and in wider popular discourses, instead of the traditional and historical term prostitution—which derives from the Latin *prostitut*, to expose publicly. The terms prostitute and prostitution are predominately used in legislation written to control, regulate, and criminalize sex work, including legislation regarding the internet. Sex work is understood to be deviant and outside of normative sexual behavior so is punished by the state. Sex work and sex workers have historically—and continue to be within digital societies—regulated and controlled through criminal laws. Hence, we find sex work in a handbook of digital criminology.

In most cases, sex work that takes place in public spaces, such as street-based sex work, which has been subject to greater policing and punishment (Blunt and Wolf, 2020). Street-based sex work remains the most stigmatized, in part to do with the fervent criminalization and the conflation with drug use. Furthermore, research suggests street-based sex workers will experience more violence than those selling sex in other spaces and in other ways (Armstrong, 2018).

This form of sex exchange continues; however, developments in digital communication technologies have transformed and expanded sex markets in three notable ways (Sanders et al., 2018; Rand, 2019). First, the emergence of new digital modalities of sex work, such as webcamming (Jones, 2015; Sanders et al., 2018). Second, the ability to pur-

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-051

chase sex or sexual activity has extended to any time and almost anywhere through the mediation of digital technologies that cross time and place. Third, facilitating this expansion and diversification of sex markets are digital platforms, acting as intermediaries between clients and workers (Rand, 2019). The pivotal and transformative role of platforms in the sale of sex has been coined the "platformatization of sex work" (Van Doorn and Velthius, 2018: 179; see Platform by Egbert). The term captures the pervasive role of platforms in the digitization of sex work, in terms of organization, regulation, and labor processes highlighting. Platforms are not merely intermediaries in the organization of sex work. This entry will define digital sex work, consider what the digitization of sex work means for sex workers regarding working practices, and enquire into the governance of digital sex work.

Digital sex work

Digital-only sex work includes the sale of sexual services mediated by technology, such as webcamming, erotic phone lines, and text services, as well as the sale of sexual content such as photographs and video reels. Webcamming, and platforms dedicated to webcam performances, have risen in popularity since they first emerged in 1996 (Jones, 2020). This section of the sex industry has become "one of the leading segments of the online sex industry in terms of revenue and numbers of customers" (Stegeman et al., 2023: 2). 'Camming' as it is popularly known, involves sex workers streaming live performances that can include flirting, flashing, stripping, autoerotic stimulation, use of sex toys, and role play (Stuart, 2022).

Currently, there are two main business models used on webcamming platforms. In the first instance, sex workers charge customers a fee per minute. In a 'group' performance, there will be several customers watching, each paying the fee set by the sex worker. In this business model, customers can request to 'go private' and pay a higher fee for one-to-one interactions. In the other business model, viewers are encouraged to 'tip' sex workers with tokens purchased from the platform. Viewers can watch for free, but sex workers commonly have a menu of sexual acts with monetary targets attached, thus encouraging viewers to 'tip' for the performance to proceed into more pornographic content (Jones, 2020; Stuart, 2022).

The internet has also changed how direct, in-person sex work is advertised, with many sex workers choosing to advertise their services on adult platforms rather than in traditional ways such as cards in telephone boxes or shop windows, standing in public spaces, and/or working in red-light districts. In the sex work literature, it is often suggested online advertising has made selling direct services safer, as sex workers can screen and report unwanted clients to the platform. This can result in the client being blocked from the site (Sanders et al., 2018). However, Hardy and Barbagallo (2021: 539) dispute this, suggesting there is an increasing availability of risky and unsafe practices because of the "visible labor oversupply" afforded by the platforms; thus giving the clients negotiating power.

As with other forms of work organized and managed via platforms, there is a distinction between sex work that takes place in-person (in the platform labor literature this is known as geographically *tethered work*) and sex work that is mediated through online platforms, or other digital technologies (known in the platform labor literature as *cloud work*) (see Woodcock and Graham, 2020 for a critical introduction to platform work). The distinction is important due to the experiences of labor and the physical and health risks associated with in-person sex work as opposed to technology-mediated sex work such as webcamming. Performing a live or pre-recorded striptease on a web camera differs significantly from having penetrative sex in an isolated hotel room or rented flat. Furthermore, the distinction is key in terms of the profit-making opportunities for the platforms (Hardy and Barbagallo, 2021).

In both cases, the platform acts as a broker between those selling services and those buying the services. Legal restrictions in most jurisdictions do not allow third parties to make money from prostitution, direct in-person sex work. However, platforms can make an income from extracting a percentage of the fee paid by the customer for digitally mediated forms of sex work. Although this varies, platforms take between 30 % and 60 % of the income generated by sex workers in each transaction (Rand, 2019). Platforms also charge for advertising and encourage sex workers to regularly upload new digital content in the form of photos and video reels. The content is freely uploaded as it is an essential part of advertising and making money as a digital sex worker (Rand, 2019). Increasingly, digital sex work is centered around the relationship with 'fans.' Digital sex workers engage in relational labor to build an online presence and engage with 'fans' to secure future work (Rand and Stegemen, 2023). Not unlike online 'influencers' whose labor relies on building audiences for their cultural production.

For many sex work scholars, the platforms are understood as extractive, acting more as third-party managers rather than merely facilitators of sexual exchanges. Early research on the camming industry suggested sex workers had greater agency over the content of digitally mediated sexual services than when working in offline markets (Jones, 2015), but as the industry has developed and been subject to greater regulation, scholars note the increase in the control over content (Stegeman, 2024; Tidenberg, 2021). For example, some platforms only allow cisgender women to sign up as webcammers or have restrictions on what words can be used (Easterbrook-Smith, 2022).

There are increasing concerns regarding the property rights of content streamed and uploaded by digital sex workers. The intellectual property rights of the content posted by sex workers are, in the main, owned by the webcam platforms, as stipulated in the platforms' terms of service (Stegeman, 2024). Due to this ownership, Stuart (2022) claims webcam platforms sell and distribute sex workers' content without renumerating them, beyond the initial fee paid by the customers and without control over where the content is posted online. Furthermore, digital sex workers are exposed to risk associated with capping (see Abuse by McAlinden and Vulnerability by Ranchordas and Beck). Capping refers to the non-consensual recording or copying of webcam perform-

ances, photographs, and videos which are then often distributed without consent on other platforms (Jones, 2020).

The digitization of sex work has offered greater degrees of control over working conditions and increasing flexibility of hours that is often appreciated by workers (Rand, 2019). Yet, digital sex workers can find themselves dependent on adult platforms more so than other platform workers because of the stigma, financial discrimination, and criminality associated with sex work (Easterbrook-Smith, 2022). This gives the platforms the power to operate in favor of profit margins rather than workers' safety and well-being. For example, the design of the platforms promote competition and price suppression and attempt to limit solidarity amongst digital sex workers (Rand and Stegeman, 2023). These concerns raised by academics and activists have not been addressed in the most recent legislation that concerns digital sex work.

Regulation and legislation of digital sex work

Sex work remains in most parts of the world, to lesser and more degrees criminalized; apart from Aotearoa (New Zealand), Belgium, and New South Wales and Northern Territories in Australia (see Global Network of Sex Work Projects (2023) for case studies and details of national legislation). National legislation regarding the internet has been slow to catch up with the rapid and dynamic changes it has brought to our daily lives. So, it is only recently we have seen nation-states actively pursuing legislation. Often the focus of this legislation is based on morality and access to sexual content. Scoular notes that during periods of economic, social, and cultural changes, prostitution laws are targeted as a site of "moral regulation" (2010: 15). Thus, legislation to control sex work and sex workers are created in an attempt to maintain the hegemonic order. In many ways, this can be seen with the concerns from many governments regarding online legislation by populist, rightwing administrations.

It is with this in mind, the following section addresses legislation enacted in the United States, under the presidency of Donald Trump. Scholars of platform governance suggest the moral and cultural values of the US have created the internet due to the concentration of tech companies originating and operating in the US (Gillespie, 2018). This dominance continues through legislation. In 2018, the US government enacted the Fight Online Sex Trafficking Act and Stop Enabling Sex Traffickers Act. FOSTA-SESTA, as it is commonly known, legislated to hold platforms responsible for "promoting or facilitating prostitution" or "knowingly assisting, facilitating or supporting sex trafficking" (Allow States and Victims to Fight Online Sex Trafficking Act of 2017). The law aims to stop trafficking for sexual exploitation in the belief that if sex markets are eradicated, all forced and coerced sex will stop. A common view held by what is sometimes known as the abolition lobby (see Mac and Smith, 2018). Sex workers in the US have widely criticized FOSTA-SESTA for being naïve at best, and more likely to be motivated by sexual morality and anti-migration ideologies. So far, there is no evidence to suggest FOSTA/SESTA has reduced sexual exploitation (Blunt and Wolf, 2020).

In the UK and Australia, legislation in the form of Online Safety Acts have seen further criminalization of 'prostitution.' There is a concern among sex workers and their allies, that these acts, including FOSTA-SESTA, silence sex workers online through over-censorship. Research in the US has shown platforms have failed to differentiate 'promoting prostitution' and 'supporting sex trafficking' from other content such as sex work activism, self-expression, humor, sex work, sex education, and so on. Platforms have tightened their terms and conditions concerning sexual content to avoid criminal sanctions under FOSTA-SESTA (Tidenberg, 2021).

Many platforms have banned and blocked individual sex worker accounts due to the legal and business risks of being shut down for "promoting prostitution" (Hacking-Hustling.org, 2023). Furthermore, platforms distance themselves from commercial sex, specifically full-service direct sex work, as seen with *Only Fans*, *Tumblr*, and *Craigslist.* The prosecution and closure of *Backpage*, a popular platform for advertising direct sex work sent a clear message to platforms that they will face criminal sanctions if the platforms advertise direct sex work (Tidenberg, 2021). Shortly following the Online Safety Act in Australia, *Switter*, an online community for sex workers shut down due to fears of defamation lawsuits. Sex worker activist organization, Hacking/Hustling refer to this as a gentrification of the internet (Blunt and Wolf, 2020).

Closing platforms, dissociation of other platforms, and the blocking of sex workers' online accounts impact sex workers' ability to participate in cultural, social, economic, and political life. This is evident in several ways. First, limiting choices on where and how to work does not eradicate sex markets, but rather marginalizes the work into more dangerous and exploitative spaces. Second, it reduces opportunities for sex workers to screen clients, having been reported as one of the key benefits of using platforms to sell sex. Third, sex workers report that their activist accounts on social media platforms have been blocked and removed (HackingHustling.org, 2023). This risks reducing opportunities for sex workers to collectively organize online and share information; thus, threatening a long history of self-advocacy and support (Easterbrook-Smith, 2022).

Digital activism

That said, the collective response to FOSTA/SESTA has seen a swell of digital activism and collective organizing by sex workers and allies. For example, shortly after FOSTA/SESTA was enacted, the campaign #SurvivorsAgainstSESTA began. The connected website and social media pages provide an online space for sex workers to disseminate information regarding payment processors, social media platforms, and crowdfunding platforms that discriminate against sex workers; how content is policed; and to organize events offline —protests, sex work benefit shows, and writing workshops. As Feldman (2014) notes, the virtual anonymity of the internet allows sex workers to 'come out' online and champion issues that impact them as sex workers. For example, hacking/hustling, "a collective of sex workers, survivors, and accomplices working at the intersection of tech and social justice to interrupt violence facilitated by technology" have organized in response to

FOSTA-SESTA. They conducted research and created resources and carried out research to "center people in the sex trades as producers of knowledge and expertise in the movement to create safety for our communities without policing" (hackinghustling.org, 2023). Hearing and listening to sex workers are vital in all research relating to sex work. Advocacy groups such as hacking/hustling, NSWP, and ESWA centre the voices and experiences of sex workers.

Conclusion

Digital societies are dynamic and ever-changing, having dramatically changed how we conduct our daily lives. New technologies develop, replacing and making redundant current technology. Lawmakers and politicians try to keep a handle on these changes and in the process further marginalize and criminalize sex workers. In attempt to address trafficking for sexual exploitation, developments in legislation can end up further criminalizing sex work. This is an ever-evolving legal, social, and cultural landscape.

The key takeaways from this chapter are:

- Sex work is a broad term to incorporate the sale of sex or sexual services in exchange for goods or money. Not all of those who exchange sex/sexual services will define themselves as a sex worker (Blunt and Wolf, 2020).
- All sex workers experience stigma, but how they negotiate and navigate stigma intersects with the context and racism, xenophobia, classism, and transphobia.
- Sex work is experienced differently depending on the person's gender, race, class, sexual orientation, place of origin, citizenship status, and other identity-related factors.
- Regretfully, the shift online has not led to further decriminalization, despite the unilateral support for decriminalization by all sex work advocacy organizations globally, international human rights organizations, and the World Health Organization.
- The legislation addressing online safety does not address the harms sex workers face by selling sex and sexual activities online, rather the law further criminalizes sex workers and further responsibilizes the platforms.

Suggested reading

Barton, B., Brents, B., and Jones, A. (2024) *Sex Work Today* New York: NYU Press.

Dewey, S., Crowhurst, I., & Izugbara, C. (eds.) (2019). *Routledge International Handbook of Sex Industry Research.* London: Routledge.

Global Network of Sex Work Projects (2023) https://www.nswp.org/

Hacking/Hustling (2023) https://hackinghustling.org/ A collective of sex workers, survivors and accomplices working at the intersection of tech and social justice to interrupt violence facilitated by technology.

Jones, A. (2020). *Camming: Money, Power, and Pleasure in the Sex Work Industry.* New York: NYU Press.

Mac, J., & Smith, M. (2018). *Revolting Prostitutes: The Fight for Sex Workers.* London: Verso Trade.

References

Allow States and Victims to Fight Online Sex Trafficking Act of 2017. Public Law 115 – 164. https://www.congress.gov/bill/115th-congress/house-bill/1865/text

Armstrong, L. (2018). Stigma, decriminalisation, and violence against street-based sex workers: Changing the narrative. *Sexualities*, 22(7 – 8), 1288 – 1308. doi: 10.1177/1363460718780216.

Blunt, D., & Wolf, A. (2020). Erased: The impact of FOSTA-SESTA and the removal of Backpage on sex workers. *Anti-Trafficking Review*, 14, 117 – 121. https://doi.org/10.14197/atr.201220148

Easterbrook-Smith, G. (2022). OnlyFans as gig-economy work: A nexus of precarity and stigma. *Porn Studies*, 10(3), 252 – 267. https://doi.org/10.1080/23268743.2022.2096682

Feldman, V. (2014). Sex work politics and the internet: Forging local and translocal political communities through the blogosphere. In S. Majic & C. Showden (eds.), *Negotiating Sex Work: Unintended Consequences of Policy and Activism* (pp. 243 – 266). Minneapolis: University of Minnesota Press.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media.* New Haven: Yale University Press.

Global Network of Sex Work Projects. (2023). https://www.nswp.org/

Hardy, K., & Barbagallo, C. (2021). Hustling the platform. *South Atlantic Quarterly*, 120(3), 533 – 551. https://doi.org/10.1215/00382876-9154898

Jones, A. (2015). Sex work in a digital era. *Sociology Compass*, 9(7), 558 – 570.

Jones, A. (2020). *Camming: Money, Power, and Pleasure in the Sex Work Industry.* New York: NYU Press.

Rand, H. M. (2019). Challenging the invisibility of sex work in digital labour politics. *Feminist Review*, 123(1), 40 – 55. https://doi.org/10.1177/0141778919879749.

Rand, H. M., & Stegeman, H. M. (2023). Navigating and resisting platform affordances: Online sex work as digital labor. *Gender, Work and Organization*, 30(6), 2102 – 2118.

Sanders, T., Scoular, J., Campbell, R., Pitcher, J., & Cunningham, S. (2018). *Internet Sex Work: Beyond the Gaze.* London: Palgrave Macmillan.

Scoular, J. (2010). What's law got to do with it? How and why law matters in the regulation of sex work, *Journal of Law and Society*, 37(1), 12 – 39.

Stegeman, H. M. (2024). Regulating and representing camming: Strict limits on acceptable content on webcam sex platforms. *New Media and Society*, 26(1), 329 – 345. https://doi.org/10.1177/14614448211059117

Stegeman, H. M., Velthuis, O., Jokubauskaitė, E., & Poell, T. (2023). Hypercategorization and hypersexualization: How webcam platforms organize performers and performances. *Sexualities*, 0(0), pp. 1 – 19. doi: 10.1177/13634607231170174.

Stuart, R. (2022). Webcam performers resisting social harms: "You're on the web masturbating… It's just about minimising the footprint". *International Journal of Gender, Sexuality and Law*, 2(1), 171 – 198. doi: 10.19164/ijgsl.v2i1.1259.

Tiidenberg, K. (2021). Sex, power and platform governance. *Porn Studies*, 8(4), 381 – 393. doi: 10.1080/23268743.2021.1974312.

Van Doorn, N., & Velthuis, O. (2018). A good hustle: The moral economy of market competition in adult webcam modeling. *Journal of Cultural Economy*, 11(3), 177 – 192. https://doi.org/10.1080/17530350.2018.1446183

Weitzer, R. (2010.) *Sex for Sale: Prostitution, Pornography, and the Sex Industry.* New York: Routledge.

Woodcock, J., & Graham, M. (2020). *The Gig Economy: A Critical Introduction.* Cambridge: Polity.

Keith Hayward

51 Smart city

Abstract: We live in the age of the smart city. This brief introductory chapter outlines some of the main features associated with this profound reconfiguration of the urban order. More specifically, it introduces criminologists to some of the many challenges posed by a world shaped by urban smartness and predicated on environmental-behavioral control.

Keywords: smart city, urban smartness, crime prevention, crime–technology nexus

Introduction

As part of the 'Saudi Vision 2030 Project,' Crown Prince Mohammed bin Salman recently greenlighted 'Line City,' a hypermodern structure 170 km long but only 200 m wide that will stretch eastwards from Saudi Arabia's Red Sea coast across the desert province of Tabuk. According to the glossy promotional renderings and virtual reality videos, Line City will accommodate up to 9 million residents on a reduced infrastructure footprint of only 34 square kilometres. It will run, it is claimed, on 100 % renewable energy, produce no pollution and, because of its unique linear design, allow every resident access to all facilities within only a five-minute walk. Described by designers as a blueprint for what 'cities of the future will look like,' Line City is the latest incarnation of the smart city paradigm—a technologically sophisticated urban area that uses digital and intelligent devices, data analytics, and innovative architectural design to overcome traditional metropolitan challenges (such as traffic congestion, crime and incivility, and environmental pollution) and enhance the life of city dwellers. For futurologists and tech company boosters, urban design concepts like Line City are a cause for celebration, a manifestation of our ability to harness technology in the pursuit of human progress and environmental sustainability. For others, the Crown Prince's grandiose and almost certainly unrealizable plans for a linear city in the desert are just the latest example of the hubris of the utopian urban planner (Rennolds and Porter, 2024; Günel, 2019), only this time filtered entirely through the lens of computational and technological advancement.

From Saudi Arabia to Songdo, Stockholm to Singapore, cities around the world are being radically transformed by developments taking place under the rubric of urban smartness. It is is hegemonic in discussions of 21st-century city planning and governance such that it is now the preeminent urban ideal of our time. But while expenditure on smart urban projects is growing rapidly (estimates suggest the global smart city market will exceed $1trillion by the end of this decade, Vidyasekar, 2013), siphoning money away from more traditional forms of public expenditure, so too are concerns about what 'living' will actually mean when urban life is ultimately planned, defined, and enforced by technological systems. In particular, critics from disciplines like urban

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-052

studies and cultural geography are worried that, by prioritizing computational prediction over-and-above the serendipitous or informal aspects of urban life, smart cities run the risk of transforming urban space into a sterile 'technological place' (Keymolen and Voorwinden, 2020), a one-dimensional, geo-fenced zone of control and coercion where the public is managed and 'responsibilized' to such an extent that individuals are reduced to unsympathetic objects of visibility and datafication (Sheehey, 2019). These and related concerns like the threat posed to privacy and liberty by intrusive surveillance (see Surveillance by Lyon), or the very real possibility that an emerging technological order will quickly fall prey to multinational corporations and private influence (Zuboff, 2019), also have obvious and very serious criminological implications. It is therefore essential that any emergent digital criminology places a critical analysis of the smart city paradigm at the very center of the enterprise.

The smart city: origins and overview

A smart city is an urban zone that uses information and communication technology (ICT), sophisticated digital data-gathering devices (such as electronic scanners, biometric sensors, and intelligent networks of connected objects and machines), and a comprehensive surveillance infrastructure to improve urban operations, manage resources, accelerate economic development, and enhance the quality of life for the zone's inhabitants (see Internet of Things by Milivojevic). At least, that's how the concept is sold. In reality, the term smart city (along with its synonym urban smartness) is better understood as a blended concept (Albino et al., 2015), combining ideas and insights from a range of different fields including computer science, cybernetics, consumerism and 'choice-based' citizenship models, artificial intelligence, management science, big data and networked informatics, environmental resilience and the 'green technology' movement, and even virtual reality and computer gaming. This broad, constitutive heritage allows the smart city paradigm to be deployed in any number of contexts to serve any number of masters. But beneath this conceptual plasticity, one (commercial) theme is always present: the accumulation of corporate profit by selling not just technology, but a utopian vision of a tech-driven, perfectly ordered, and functioning city. To understand why this is the case, one needs to know something of the historical origins of the smart city concept.

In the 1990s, the growth of computing and the subsequent massive expansion of the internet triggered a rash of new initiatives that sought to use ICT to solve urban problems. At this point, the name smart city was just one of many terms—information city, cyber city, intelligent city, wired city, knowledge city, etc.—that were being used to describe developments in this area. In these early iterations, themes like 'competitiveness' and 'economic efficiency' were often balanced alongside other key elements, such as sustainability, citizen empowerment, e-governance, or bottom-up community engagement (Montes, 2020). However, this weighting equivalence quickly tilted in favor of corporate interest. In 2008, after a decade of financial losses, the world's preeminent

computer systems company, IBM, was staring down the barrel of a global financial crisis and desperately needed something to revive its flagging revenue streams. It found it in then CEO Sam Palmisano's much publicized talk 'A Smarter Planet: The Next Leadership Agenda.' A year later they launched a multi-million-dollar 'Smarter Cities' marketing initiative, officially registering that same term as a trademark in 2011. According to Söderström and colleagues (2014), from here, IBM set about defining and promoting urban smartness through a series of publications, promotional tradeshows, and high-profile conference events. It proved an extremely successful strategy. By solidifying their position as the key stakeholder in the field, IBM were perfectly placed to first develop and then exploit the global smart city market. In particular, they discursively constructed the smart city as a place of possibility, an *open* space for playful corporate experimentation that would simultaneously enhance urban 'creativity' and encourage inward financial investment. Other companies, most notably Cisco, Siemens, Microsoft, General Electric, and later, Google's Sidewalk Labs, quickly followed suit, presenting a vision of the smart city premised on a consistent set of themes—cost efficiency, security, environmental sustainability, algorithmic synchronicity, seamless digital connectivity, etc.—all of which would supposedly operate in a neutral, non-ideological fashion. The reality, of course, was/is entirely different. Behind the sheen of openness and creative innovation was a deep commitment to profiteering and global neoliberalism (see Sadowski and Bendor, 2019), something predicted by Hollands (2008) in an early critical article that portrayed the then embryonic smart city as just another form of urban entrepreneurialism. But although Hollands was right to see the smart city as a variation/extension of the marketable "creative city" rubric, at this point he could only sense how smartness providers would come to dominate the market and the surrounding discourse. By continuing to establish and promote a corporate version of smartness, these companies ensured that, today, no meaningful alternative (non-neoliberal) smart city model exists. It is for this reason that smart cities and the various applications and developments associated with urban smartness have been the subject of both academic criticism (see e.g., Greenfield, 2013; Rosati and Conti, 2016), and more recently nascent forms of protest and urban resistance designed to check the advance of techno-imperialism.

Criticisms, criminological concerns and future scenarios

The underlying goal of smart city planners and designers is to create urban zones where every action, interaction, incident, and exchange, is surveilled, recorded, datafied, and networked. In-and-of-themselves, such practices could theoretically be neutral. For example, in countries like Denmark, Switzerland, and Sweden, where the public's trust in governments and municipal authorities is historically high, smart city initiatives like traffic flow monitoring, public transport patterning, and recycling man-

agement/processing have been rolled out without incident, with the majority of residents benefiting from these developments. However, elsewhere in the world, it's often the case that, not only do some of the technologies associated with smart cities fail to address the underlying tensions and inequalities that shape the contemporary urban order, but in some cases, they actually make the situation worse (see Creemers, 2018; Mozur, 2019). For example, in the city state of Singapore, it is claimed that the ruling People's Action Party are using smart technology as a tool for suppressing anti-government politics—so much so in fact that the Singaporean regime has been described by one commentator as a "quasi-autocratic technocracy" (see Guest, 2021). It is this aspect of urban smartness that should concern criminologists of all stripes and thus in the brief space afforded me here I set out some (but certainly not all) of the unresolved concerns and criticisms that continue to swirl around the smart city paradigm.

The first and perhaps most pressing issue is the emergence of what one might call the surveillance–smartness nexus. There is little doubt that the continued merger of sophisticated data-gathering devices (such as biometric sensors and intelligent networks of connected objects and machines) and totalizing surveillance infrastructure (e.g., military-grade Wide Angle Motion Imagery (WAMI) reconnaissance platforms (Michel, 2019) and city-wide security assemblages known as Domain Awareness Centres (Sadowski, 2019)), will bring about a reduction in the type of acquisitive street crime that frequently blights urban environments. But while we all want to live in safer, securer cities, there are growing fears that this type of techno-overwatch may come at a significant cost. In particular, it is argued that all-seeing, data-aggregating systems created from a union of surveillance and artificial intelligence will transform cities into sterile, privacy-eroding zones of control and coercion (Krivý, 2018).

Second and relatedly is the threat posed to liberty and democracy by the very real possibility that the only smart city future is one shaped and directed by multinational capitalism and corporate discourse (Sadowski, 2020). This view, shared by many critical and cultural criminologists, is encapsulated by Brunilda Pali and Marc Schuilenburg (2020) in the following quote from a recent article about fear and fantasy in the smart city: "If, however, we gaze beneath the clichés and rhetoric, the smart city appears as a 'naked king'—a commercial construct designed to sell a corporate vision of capital accumulation, which necessitates different types of surveillance to achieve it."

A third issue concerns the use of smart technology as a criminal or adversarial vector (see Cybercrime by Holt and Holt). Most obviously, by developing advanced systems of digital/biometric surveillance and pre-emption in an attempt to make 'traditional' crime forms impossible, are we simply exposing society to a new and often unanticipated array of 'technocrimes' (Steinmetz and Nobles, 2017)? These would include everything from crimes *with*, *against*, and *by* artificial intelligence (see Hayward and Maas, 2021) to the use of bio-tech body implants and transhuman ('wetware') biohacks to overcome security protocols (Goffette, 2017). Similarly, one must also consider the emerging phenomenon of the so-called 'Techlash' (Hayward, 2025). While urban smartness offers up new possibilities of situational-behavioral control, it also opens the door

to an alternative, more confrontational future that includes urban resistance and criminal destruction designed to counteract the rise of an all-seeing surveillance state (see e. g., Byrne and Davis, 2020; Luque-Ayala and Marvin 2020). This focus on political grievance and illegal actions against technology (as a material symbol of spatial marginalization or oppressive governance) rather than the more traditional criminological focus on acquisitive types of criminality/exploitation is likely to become an ever-more pressing problem for criminology—not least because anti-technology 'cultures of resistance' are already starting to crystallize (see e. g., Bonini and Treré 2023; Smith and O'Malley, 2017). In sum, by accelerating into a world of smartness, might we simply be exposing society to a largely unanticipated series of risks and system failures that ultimately could bring down the fabric of the liberal order (see Bridle, 2018)?

Finally, and here I am projecting even further into the future, concerns have also been raised that the constructivist smart city of the mid-21st century will result ultimately in the "progressive cybernetisation of urban life" (Krivý, 2018). A starting position here is Jennifer Gabrys' (2014) observation that, while the smart city is fundamentally an environmental phenomenon, it also involves an active social function in the way it transforms urban citizens into 'citizen sensors.' This process will take many forms. For example, the current obsession among tech companies and certain governments with epidermal and textile-embedded 'wearables' (digital activity recorders/transponders that are used to track humans and produce detailed spatial and physiological data trails) could potentially transform the urbanite into little more than an automated behavioral node. A less overt but potentially even more worrying development relates to the shifting nature of consciousness and cognition of urbanites. Here the argument is that, as smart citizens come to rely ever-more on the network to maintain them in "a state of perpetual safety and total security" the experienced division between humans and the cybernetic machine could potentially dissolve entirely (McGuire, 2018: 8). Through processes such as "inattentional blindness" and the "atrophying of our critical faculties" human agency is "filleted out," to use McGuire's terminology. As a result, smart citizens would lose not only the ability to govern themselves but more importantly, *the will to govern themselves* as the intensity of data-flow(s) between humans and non-human systems grows ever greater; a relational dynamic that will likely further diminish social resilience and exacerbate passive infantilization (Hayward, 2024).

Conclusion

As society becomes ever-more reliant upon smart technology, including sensor-enabled interactive devices and interconnected data architecture, we accelerate further into a world where the greater the connectivity, the greater the potential exposure to criminal attacks. However, as it stands, the discipline of criminology, despite a long and productive interest in the relationship between urban space and criminality (Hayward, 2004), has largely ignored the serious problems posed by a world increasingly shaped

by urban smartness and predicated on environmental-behavioral control. In particular, many of criminology's existing methods and causal theories, based as they are on utilitarian models of behavior and mid 20th-century notions of deviance, are often not appropriate for a world in which traditional notions of social control are being superseded by dissipative systems that curate individuals within digitally regulated and restricted terrains. This being the case, then, there is an urgent need for criminology to evolve and adapt so that it's more appropriately equipped to deal with the many challenges posed by smart cities and the various other societal realignments brought about by intelligent technology and digitalization.

Suggested reading

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3 – 21.

Bridle, J. (2018). *New Dark Age.* London: Verso Books.

Greenfield, A. (2013) *Against the Smart City*, New York: Do Projects.

Hayward, K. J. (2021). Five smart city futures: A criminological analysis of urban intelligence. In M. Schuilenburg & R. Peeters (eds.), *The Algorithmic Society* (pp. 172 – 192). London: Routledge.

Sadowski, J. (2020). *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World*, Boston: MIT Press.

References

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3 – 21.

Bridle, J. (2018). *New Dark Age.* London: Verso Books.

Bonini, T. and Treré, E. (2023) Algorithms of Resistance. Boston: MIT Press.

Byrne, R., & Davis, M. C. (2020). Protest tech Hong Kong. *The Wilson Quarterly.* http://wq.proof.press/quarterly/the-power-of-protest/protest-tech-hong-kong/

Creemers, R. (2018). *China's Social Credit System: An Evolving Practice of Control.* Rochester, NY: Social Science Research Network.

Gabrys, J. (2014). Programming environments: Environmentality and citizen sensing in the smart city. *Environment and Planning D: Society and Space*, 32(1), 30 – 48.

Goffette, J. (2017). Technology, body and human enhancement. In M. R. McGuire and T. J. Holt (eds.), *The Routledge Handbook of Technology, Crime and Justice* (pp. 654 – 672). London: Routledge.

Greenfield, A. (2013). *Against the Smart City.* New York: Do Projects.

Guest, P. (2021). Singapore's tech-utopia dream is turning into a surveillance state nightmare. restofworld.org, 16th November Available at: https://restofworld.org/2021/singapores-tech-utopia-dream-is-turning-into-a-surveillance-state-nightmare/

Günel, G. (2019) Spaceship in the Desert, North Caroline University Press.

Hayward, K. J. (2004). *City Limits: Crime, Consumer Culture and the Urban Experience.* London: Cavendish.

Hayward, K. J. (2024). *Infantilised: How Our Culture Killed Adulthood.* London: Little Brown.

Hayward, K. J. (2025). The Smart City as adversarial vector: Some criminological reflections on the 'preventative technological imaginary'. In F. Donnelly & T. Schwarze (eds.), *Security and Space.* Bristol: Bristol University Press.

Hayward, K. J., & Maas, M. (2021). Artificial Intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, 17(2), 209 – 233.

Hollands, R. (2008). Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? *City*, 12(3), 303 – 320

Keymolen, E. and Voorwinden, A. (2020). Can we negotiate? Trust and the rule of law in the smart city paradigm. *International Review of Law, Computers, and Technology*, 34(3), 233 – 253.

Krivý, M. (2018). Towards a critique of cybernetic urbanism: The smart city and the society of control. *Planning Theory*, 17(1), 8 – 30.

Luque-Ayala, A. and Mariv, S. (2020) Urban Opeating Systems. Boston: MIT Press.

McGuire, M. (2018). Beyond flatland: When smart cities make stupid citizens. *City, Territory and Architecture*, 5, Article Number 22.

Michel, A. H. (2019). *Eyes in the Sky: The Secret Rise of Gorgon Stare and How It Will Watch Us All.* Boston, MA: HMH Books.

Montes, J. (2020). A historical view of smart cities: Definitions, features and tipping points. Available at: http://dx.doi.org/10.2139/ssrn.3637617

Mozur, P. (2019). One month, 500,000 face scans: How China is using A.I.to profile a minority', *The New York Times*, 14 April. Available at: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

Pali, B., & Schuilenburg, M. (2020). Fear and fantasy in the smart city. *Critical Criminology*, 28, 775 – 788.

Rennolds, N., & Porter, T. (2024). Saudi Arabia's futuristic $500 billion Neom megacity just got a lot smaller... for now, report says. businessinsider.com, 6 April.

Rosati, U., & Conti, S. (2016). What is a smart city project? An urban model or a corporate business plan? *Procedia: Social and Behavioral Sciences*, 223, 968 – 973.

Sadowski, J. (2019). The captured city: the "smart city" makes infrastructure and surveillance indistinguishable. realifemag.com, November 12.

Sadowski, J. (2020). *Too Smart: How Digital Capitalism is Extracting Data, Controlling Our Lives, and Taking Over the World.* Boston: MIT Press.

Sadowski, J., & Bendor, R. (2019). Selling smartness: Corporate narratives and the smart city as a technological imaginary. *Science, Technology, and Human Values*, 44(3), 540 – 563.

Sheehey, B. (2019). Algorithmic paranoia: The temporal governmentality of predictive policing. *Ethics and Information Technology*, 21(1), 49 – 58.

Smith, G., & O'Malley, P. (2017). Driving politics: Data-driven governance and resistance. *British Journal of Criminology*, 57(2), 275 – 298.

Söderström, O., Paasche, T., & Klauser, F. (2014). Smart cities as corporate storytelling. *City*, 18(3), 307 – 320.

Steinmetz, K. F., & Nobles, M. R. (2017). *Technocrime and Criminological Theory.* London: Routledge.

Vidyasekar, A. D. (2013). *Strategic opportunity analysis of the global smart city market: Smart city market is likely to be worth a cumulative $1.565 trillion by 2020.* Frost & Sullivan. Retrieved January 2024, from http://www.frost.com/sublib/display-report.do?id=M920-01-00-00-00.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: Public Affairs.

Mirjam Abigail Twigt

52 Social media

Abstract: This chapter discusses three key developments sparked by social media relevant to criminologists: social media as enablers of harm, surveillance, and manipulation. Concerns that social media for crime and crime control purposes contribute to discrimination relate to the general agreement that social media and social media data tend to reinforce pre-existing structural inequalities and power inequalities. Another concern relates to the pace at which companies develop means to gather and use social media intelligence, further hindering capacities for meaningful regulation and oversight. The connections that social media engender however also provide additional means to collectively respond to data harms and to foster new approaches for anti-surveillance and social justice.

Keywords: platforms, profiles, sociality, datafication, privacy work

Social media have become integral to much of everyday life. Most of us use social media platforms to communicate with family and friends, share content and to interact with events in the world. "Social media are by definition social" (Marwick, 2023: 6). Much like offline interactions, the online connections that social media enable are deeply embedded in material realities and social relations. What is new(er) about social media is the ways in which they enhance interactions across spatial and temporal boundaries.

The term 'social media' comes from popular culture grouping together different interactive Web 2.0 internet-based applications that draw on user-generated content (Fuchs, 2017). The first social media sites gained popularity in the 1990s. Some platforms such as Myspace (2003) have waned, whereas others like YouTube (2005, owned by Google) and Facebook (2004) remain successful. More social media platforms continue to emerge, such as TikTok and its Chinese counterpart Douyin, Twitter/X, Viber, WeChat, and SnapChat, which exist next to other popular platforms owned by Meta (formerly Facebook. Inc.) like Instagram and WhatsApp. As smartphones and access to the internet have become less expensive, social media gained further popularity and an ubiquitous global presence.

Social media platforms can be considered as additional places for people to socialize, with different degrees of privacy and group size. They have engendered more scales for sociality as they offer additional means to sustain connections and to build new networks, to establish romantic and/or sexual encounters and to deal with experiences of isolation and exclusion—just to name a few (Miller et al., 2016). Most social media platforms expect users to create service-specific and highly personalized profiles. This allows for connecting and sharing content with other individuals and groups through one-to-one, one-to-many and/or many-to-many exchanges. Many users combine different social media sites.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-053

The socio-technical properties of social media profoundly reshape social relations and politics (Dencik et al., 2022). Amongst other things, they interact with and alter practices and understandings concerning crime, punishment, and control. Online interactions allow additional ways to engage in harmful or illicit acts, in policing and resisting control, not least because extracted social media data can be used for surveillance and manipulation purposes. Processes of extracting and using social media data tend to be unequally distributed across social relations and the global political economy with much risk of further structural discrimination (Couldry and Mejias, 2019). Terms such as 'data universalism' and other manifestations of technological determinism relating to social media easily gloss over material and cultural specificities. In the Global North and across the many Souths there are manifold ways in which people navigate the respective possibilities and restrictions of different social media platforms (Milan and Treré, 2019).

Social media as enablers for new forms of harm and/or criminal acts

Social media platforms allow for new criminal and hurtful practices and for old(er) practices to take place on them. Much has been written on their potentials for amplifying old and new forms of hate, abuse, and discrimination (Noble and Tynes, 2016). For instance, social media platforms have frequently been used to incite racist hatred, using anonymity or fake identities, and by employing seemingly benign practices like the use of memes, GIFs, and emojis (Matamoros-Fernández and Farkas, 2022). The affordances of different social media allow for new ways to organize crime transnationally: for instance, online drug traders in Scandinavia combine the use of the dark web and encrypted one-to-one communication, with public-facing social media profiles that are used for assessing risks and build trust (Bakken, 2022).

Much of social media's potentials, also for inflicting harm, relate to their design, which encourages sharing. Their design and policies enable practices of sharing and coordinating dissatisfaction with society, libidinal desires, and rage online (Paris, 2021). Social media are also prime channels for sharing and spreading false media content (Venturini, 2019). Many social problems that emerge online—such as harassment or cyberbullying—have roots in offline social dynamics. But social media platforms allow for additional ways to know about someone's past or a person's private information, and to direct hurtful comments. There are specific concerns on the influence of social media on young people's mental health (Angel and boyd, 2024).

An important difference between social media and other interpersonal communication, such as letters and landline phones, is that information shared in a closed online community or in a confidential setting can easily—intentionally or unintentionally—leak into more public channels and to contexts for which it was not intended (Chun, 2017). Whereas this can happen on a personal level, data breaches and hacks also often

occur on a larger scale: Facebook and Twitter/X are among the world's largest social media companies through which personal data of millions of users have become exposed, information that could be misused for reputational damage, fraud, and identity theft (Information is beautiful, 2024; see ID Theft by Langford, Wærstad, and Svensson).

The exposure to online hatred and other social media-related harms is not equally divided. People of color, women, and LGBTQI+-individuals are more often targeted. The societal repercussions tend also to be greater for them, not least because legal systems are still catching up on how to address social media-facilitated harms and many societal and material barriers for legal recourse exist (Citron, 2016).

Social media as enablers of surveillance and control

Surveillance—the systematic collection of personal information to manage groups—is not new, but the rise of social media adds new dimensions (see Surveillance by Lyon). Firstly, surveillance is deeply embedded in social media infrastructures and business models. Many social media platforms are intricately linked to what Shoshana Zuboff has critically termed *surveillance capitalism* (Zuboff, 2019). Such platforms commodify vast amounts of behavioral data generated through users' engagement. Infrastructures that track user activities and quantify these as data-points create capital, as millions of social media messages a day are analyzed to gain extensive profiles of users. These profiles are subsequently sold, directly or via data brokers, to advertisers, to security officers who want to control a particular target group, or as training data for the development of Artificial Intelligence (AI) (see Artificial Intelligence by Van Brakel). Vice versa, machine learning tools are also used to analyze social media interactions (Amoore, 2020).

The availability of information on social media also changes governmental surveillance. Using social media data for surveillance and policing purposes does not necessarily require advanced techniques. Research shows how UK police officers would manually search the social media presence of targeted persons, inadvertently also collecting detailed information of non-targeted persons (Fussey and Sandhu, 2022). Fake social media profiles can be set up to gather digital evidence, as has been the case in several countries where queerness is criminalized (Rigot, 2022).

Despite their potential for discrimination, data-driven systems have increasingly become integrated into governmental processes. Seemingly mundane data—clicks and likes, time spent on a site, the music preferences—are collected and put together based on shared features and common behavioral patterns the use of which can be of severe consequence: they can influence one's credit score or risk assessments conducted by the police (Redden, 2022: 70 – 71). The widespread experimentation with social media data by public institutions ranging from The Netherlands, the US, the UK, and India on areas such as predictive policing, welfare provision, and fraud prevention have had many harmful effects, such as extended surveillance, stigmatization, wrong-

ful accusation, detention, and deportation. Especially people already marginalized have been adversely affected by such practices (Redden, 2022).

Yet a different aspect are the predictions drawn from data aggregates, which are used for policing activities, as well as military purposes. For example, techniques and tools for facial recognition, sentiment analyses, mappings of suspicious relations or events draw on social media data (Trottier, 2014). Companies offer technologies for intercepting 'real time' communication on social media or for accessing a person's social media information. What is more, digital data extraction devices and tools to restore deleted social media data are not only appropriated by dictatorial regimes, allowing them to take action against dissent (Loewenstein, 2023). Several EU member states also use them to search social media traces of people seeking asylum (Josipovic, 2023). This is but one example of how social media data are used for migration and border control purposes, despite ample questions about the accuracy and the legality of such procedures (Oliveira Martins et al., 2022).

Social media as enablers of polarization and/or manipulation

Surveillance capitalism has close links to the 'attention economy' (Celis Bueno, 2017). The longer one engages with a specific social media platform, the more data can be gathered. What social media users see on their feeds is curated by algorithms (see Algorithm by Leese), for instance through the push of 'viral' content and tailored predictions based on what you and people like you 'liked.' The algorithmic selection of social media contents has thus been criticized for creating echo chambers (Jamieson and Cappella, 2008) or filter bubbles (Pariser, 2011), by making social media encounters with people with opposing views more limited and making news-intake more one-sided. This criticism has also triggered new developments in curating contents. A systematic review of existing research on social media and polarization suggests a more complex picture, with a formative role for self-selection and more vocal elites (Ross Arguedas et al., 2022).

Other sources of manipulation are bots and trolls used to spread and amplify false and/or dangerous content (Woolley and Howard, 2016), but also to set up campaigns for targeting specific user groups. The Cambridge Analytica Scandal of 2018, for example, revealed how the data of more than 80 million Facebook users was used to target particular demographic groups with tailored messages to steer voters' decision making (Redden, 2022: 67). As a response to such issues, governmental pressures and reputational concerns have stimulated social media platforms to curtail media manipulation and other social media-related harms, for instance via content moderation. Here, automated approaches are combined with the labor of content moderators to filter out violent and/or inappropriate social media content and distinguish false accounts.

Social media companies, too, have an interest in addressing manipulation and polarization, but their initiatives, too, create social repercussions. As profit-oriented businesses, social media companies are concerned with not offending the majority of their users and influential nations. When taking actions, social media platforms owners, however, tend to draw on their own normative frameworks. There is power in determining what and who causes harm, as such definitions can easily be experienced as punitive and selective (Myers West, 2018). Social-media platforms, then, also exercise normative power: Loewenstein (2023), for example, has documented how social media platforms censor, suspend, and make content disappear among already marginalized groups.

Enacting concrete policymaking and politics

Social media developments occur quickly, whereas data regulation ecologies tend to lag behind. Concerns for surveillance and other social media-related harms have been translated into regulatory frameworks for responsible data handling such as the EU's General Data Protection (GDPR). Questions remain about their scope and enforceability (Hintz, 2022). And with their narrow emphasis on the protection of personal data, these frameworks draw on the ontological premise of 'hyperindividualism' (Bigo et al., 2019: 5). The individual social media-user is required to protect themselves through technological self-defense. Meanwhile participating in an online social life on the most widely used services and platforms already requires agreeing to comprehensive data collection (Hintz, 2022).

Many social media users engage in 'privacy work' via a wide array of practices such as obfuscating personal information, self-censoring, using code language, encryption, and virtual private networks (VPNs) (Marwick, 2023). But approaching a desired level of privacy—clear boundaries between self and others and between what is public and private—is difficult, if not impossible, to sustain. Even if people fully abstain from social media, they cannot control what others share about them online, or if big data technologies make inferences about their behavior in other ways. Much of the value of social media data relates to their relational value. Data policies and legislation that focus solely on the individual data subject fail to address the relational value of social media data, and how this reshapes crime and control within societies.

The pace at which possibilities for social media intelligence gathering develop increasingly outstrips capacities for meaningful regulation and oversight. At the same time, much of the difficulties for meaningful regulation of social media is characterized by inherent contradictions within legal and regulatory frameworks. In addition to the GDPR, other EU legislation such as the Digital Markets Act (DMA) and the Digital Services Act (DSA) are developed to regulate and curtail the power of social media platforms (Meaker, 2024). These, however, coexist with the EU's ongoing emphasis on 'innovation' and specific understandings of national 'security' (Hintz, 2022).

Corporate initiatives to implement data ethics have had varying degrees of success, but these largely rely on the goodwill of the industry with little space for governmental entities to interfere. Rather, their suggestions for technological solutions (more data, algorithmic sophistication) for socio-political problems suggests that much work is ahead for civil society actors, tech workers, and activists. Among other things, this will consist of activities geared to bring forward the harms and violations brought about by social media platforms, to reappropriate decentralized social media interactions for building communities, and to question data extraction practices by social media platforms on a more profound level (Dencik et al., 2022).

Conclusion

This contribution shared a critical account of the ways in which social media interact with and have become an element of crime and crime control. The potentials of social media were initially idealized as democratizing and leveling. By now, there is general agreement that social media reinforce already existing unequal power differentials in gender, class, race, legal status, and sexuality (Chun, 2017). In extension, this raises concerns about the ways in which the use of social media for crime and crime control can further entrench structural inequalities and contribute to discrimination.

The connections that social media engender, however, also provide additional means to collectively respond to these developments and to identify new forms of data justice (see Data Justice by Redden). Social media are also means for sousveillance, the redirection of surveillance toward those traditionally in position of power. They allow to capture, share, and gain attention for injustices, such as police violence (Goldsmith, 2015), for creating counter-narratives, for establishing safe passages for people on the move (Milijovic, 2018) and for strengthening social movements through narrative capacities (Tufekci, 2017). Rather than moving away from social media, there is value in moving toward data justice approaches that further bridge anti-surveillance and social justice activism (Dencik et al., 2022).

Suggested reading

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data Justice.* London: SAGE.

Fussey, P., & Sandhu, A. (2022). Surveillance arbitration in the era of digital policing. *Theoretical Criminology*, 26(1), 3 – 22.

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech and social media: A systematic review and critique. *Television & New Media*, 22(2), 205 – 224.

Marwick, A. (2023). *The Private is Political: Networked Privacy and Social Media.* New Haven, CT: Yale University Press.

References

Angel, M., & boyd, d. (2024). Techno-legal solutionism: Regulating children's online safety in the United States. In *CSLAW'24: 3rd ACM Computer Science and Law Symposium, March 12 – 13, Boston, MA.*

Amoore, L. (2020). *Cloud Ethics: Algorithms and Attributes for Ourselves and Others.* Durham, NC: Duke University Press.

Bakken, S. A. (2021). Drug dealers gone digital: Using signalling theory to analyse criminal online personas and trust. *Global Crime*, 22(1), 51 – 73.

Bigo, D., Isin, E., & Ruppert, E. (2019). *Data Politics. Worlds, Subject, Rights.* Abingdon: Routledge.

Celis Bueno, C. (2017). *The Attention Economy: Labour, Time and Power in Cognitive Capitalism.* London and New York: Rowman & Littlefield.

Chun, W. H. K. (2017). *Updating to Remain the Same Habitual New Media.* Cambridge, MA: the MIT Press.

Citron, D. K. (2016). *Hate Crimes in Cyberspace* (Reprint ed.). Cambridge, MA: Harvard University Press.

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, 20(4), 336 – 349.

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data Justice.* London: SAGE.

Fuchs, C. (2017). *Social Media: A Critical Introduction.* London: SAGE.

Fussey, P., & Sandhu, A. (2022). Surveillance arbitration in the era of digital policing. *Theoretical Criminology*, 26(1), 3 – 22.

Goldsmith, A. J. (2015). Disgracebook policing: Social media and the rise of police indiscretion. *Policing and Society*, 25(3), 249 – 267.

Hintz, A. (2022). Data and citizenship. In L. Dencik, A. Hintz, J. Redden, & E. Treré (eds.), *Data Justice* (pp. 73 – 88). London: Sage.

Information is beautiful. (2024). *World's Biggest Data Breaches & Hack.* Updated January 2024. Retrieved from: https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

Jamieson, K. H., & Capella, J. N. (2008). *Echo Chamber: Rush Limbough and the Conservative Media Establishment.* Oxford and New York: Oxford University Press.

Josipovic, I. (2023). What can data justice mean for asylum governance? The case of smartphone data extraction in Germany. *Journal of Refugee Studies*, 36(3), 534 – 551.

Loewenstein, A. (2023). *The Palestine Laboratory. How Israel Exports the Technology of Occupation around the world.* London: Verso.

Marwick, A. (2023) *The Private is Political: Networked Privacy and Social Media.* New Haven, CT: Yale University Press.

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech and social media: A systematic review and critique. *Television & New Media*, 22(2), 205 – 224.

Meaker, M. (2024). Europe is breaking open the empires of big tech. *Wired.* 6 March, Retrieved from: https://www.wired.com/story/europe-dma-breaking-open-big-tech/.

Milan, S., & Treré, E. (2020). Big data from the South(s): Beyond data universalism. *Television & New Media*, 20(4), 319 – 335.

Milivojevic, S. (2019). 'Stealing the fire', 2.0 style? Technology, the pursuit of mobility, social memory and de-securitization of migration. *Theoretical Criminology*, 23(2), 211 – 227.

Miller, D., Costa, E., Haynes, N., McDonald, T., Nicolescu, R., Sinanan, J., Spyer, J., Venkatraman, S., & Wang, X. (2016). *How the World Changes Social Media.* London: UCL Press.

Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society,* 20(11), 4366 – 4383.

Noble, S., & Tynes, B. (2016). *The Intersectional Internet: Race, Sex, Class, and Culture Online.* New York: Peter Lang Inc.

Oliveira Martins, B., Lidén, K., & Gabrielsen Jumbert, M. (2022). Border security and the digitalisation of sovereignty: Insights from EU borderwork. *European Security*, 31(3), 475 – 494.

Paris, B. (2021). Configuring fakes: Digitized bodies, the politics of evidence and agency. *Social Media & Society*, 7(4), 1 – 13.

Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You.* London: Viking.

Redden, J. (2022). Data harms. In L. Dencik, A. Hintz, J. Redden, & E. Treré (eds.), *Data Justice* (pp. 59 – 72). London: Sage.

Rigot, A. (2022). *Digital Crime Scenes: The role of digital evidence in the persecution of LGBTQ people in Egypt, Lebanon and Tunisia.* Berkman Klein Center for Internet & Society at Harvard University.

Ross Arguedas, A., Robertson, C., Fletcher, R., & Nielsen, R. (2022). *Echo Chambers, Filter Bubbles and Polarisation: A Literature Review.* Oxford: Oxford University.

Trottier, D. (2014). Vigilantism and power users. Police and user-led investigations on social media. In D. Trottier & C. Fuchs (eds.), *Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube* (pp. 209 – 222). Abingdon: Routledge.

Tukefci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest.* New Haven, CT: Yale University Press.

Venturini, T. (2019). From fake to junk news: The data politics of online virality. In D. Bigo, E. Isin, & E. Ruppert (eds.), *Data Politics. Worlds, Subjects, Rights* (pp. 123 – 144). London: Routledge.

Woolley, S., & Howard, P. (2016). Automation, algorithms and politics|political communication, computational propaganda, and autonomous agents – Introduction. *International Journal of Communication*, 10(9), 4882 – 4890.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* London: Profile Books.

David Lyon

53 Surveillance

Abstract: Surveillance is a central concept for understanding contemporary societies, both locally and globally. Among its many uses are policing and criminology, although many characteristics of today's surveillance are shared in a digital environment. One such is that current dependence on algorithms may jeopardize just and fair practices rather than enhancing human flourishing.

Keywords: visibility, social sorting, data justice

Introduction

The concept of surveillance is central to a contemporary understanding of the digital world, including digital criminology. However, unlike some other concepts used in this context, the word surveillance is more than two hundred years old and thus has seen major social, political, and technical changes that have prompted shifts in its meaning. From being a concept that once spoke primarily of "close observation, especially of a suspected spy or criminal" (OED, 2011), in the 21st century it acquired the sense of encompassing a whole political-economic order as 'surveillance capitalism' (Mosco, 2014; Zuboff, 2015, 2019). Surveillance now speaks of an infrastructural condition, increasingly, on a global scale. In between, the concept took on varied meanings, depending on its use in differing contexts of administrative, military, policing, epidemiological, workplace, and other areas. In each, the word was both a technical term for specific activities and, from the 1970s, a concept increasingly imbued with meaning first from computing and then from the expanding digital realm (see Digital by Wernimont).

The concept of surveillance relates to practices of 'watching over,' that have developed especially in modern, Western times, aided increasingly by mechanical and digital technologies. Surveillance here refers primarily to the human world but is frequently imbricated with the non-human and with technology. The concept of surveillance is distinguished by its associations with power and resistance, and by the varying kinds of meaning-making that accompany its spread. It is a much-contested critical concept in that its meaning is not settled in common use, and it is often debated in the context of political disputes, including those of crime and policing, both locally and globally.

In what follows, I offer a definition of surveillance relating to a range of social practices and note how it is distinct from other concepts, such as monitoring or spying. I then show how the concept has evolved through four stages: observation, sorting, digi-

Disclaimers and Acknowledgments: This text is an adapted version of the previously published article: Lyon, D. (2022). Surveillance. *Internet Policy Review*, 11(4): n.p.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-054

tization, and dataveillance. Each stage represents a progressive shift further away from direct inter-personal watching and towards 'making visible,' first through organization and then machines, through to digital data (see Datafication by Chan). This prompts a discussion of the multidisciplinarity of the concept and finally to a brief survey of its analytical and practical value, as well as to the possible futures of the concept of surveillance.

The surveillance concept in context

Definition and development

The concept of surveillance as a social practice may be defined as "the focused, systematic and routine attention to personal details for the purposes of influence, management, protection or direction" (Lyon, 2007: 14). The concept points to both practices and purposes. Many qualifications are needed to fill out this definition. For example, this definition refers to 'attention to personal details,' thus allowing a stretch from 'watching' to listening and other kinds of 'attention,' including those enabled by electronic means.

The mention of 'electronic means' also hints that the simple 'watching over' of, say, a worker by an employer, is today much more subtle. Surveillance now "makes visible" (Taylor, 2017: 4) through many means, especially by data collection, analysis, interpretation, and action. Moreover, the 'making visible' achieved by surveillance might occur without any deliberate operator attention to, or awareness on the part of, particular people. Personal profiles may be constructed from disparate data, gleaned from consumer behavior and from a myriad of other apparently random sources.

Surveillance, then, is a modern concept, used in English since the 19th century as a loan-word from the French; *sur-* 'over' and *veiller* 'watch,' which both come from the Latin, *vigilare*, to keep watch. Spanish reflects this in *la vigilancia*, and *Überwachung* gives the same sense in German. Surveillance may be viewed as appropriate vigilance, to protect society from risks of attack, disease, crime, or corruption. Indeed, it may be considered as protective of freedom and liberty, as much as it is about care as control (Rule, 1974; Lyon, 1994; Taylor, 2020).

The use of the concept of surveillance, including its adverse aspects since the 19th century, is no accident. This was a period when industrial capitalism came into its own, involving new modes of organization and governance, both within emerging national and colonial governments and in new forms of economic life, like in production and consumption. From the first such usage, while direct perception was never abandoned, the technologies of surveillance were also important, entailing as they do, ways of enhancing first vision, then hearing and eventually, memory (Lauer, 2011). For example, improved lighting on the streets of Paris, to enhance visibility, was a policing priority in 1668 (Tucker, 2017). In the 1890s, San Francisco newspapers complained about tele-

phone operators listening-in on conversations (Lauer, 2011: 577), a practice soon followed by others rather than just operators. And while Thomas Edison promoted the surveillance use of his phonograph as a way of enhancing memory, in the 1880s, Edward Higgs notes that in Europe the state collection and thus 'memory' of citizen data—not only for 'control'—can be traced to the 1500s (Higgs, 2004).

However, from the mid-20th century on, surveillance itself was increasingly construed as a threat to freedom and liberty, not only when it was used to buttress Nazism and authoritarian communism but also—especially in the writings of George Orwell (1949)—in Western democracies. This negative connotation of the concept, including the *control* of the watched by watchers, is the source of much social criticism. However, some argue, the latter is not a necessary connotation (e.g., Andrejevic and Selwyn, 2022; Lyon, 2007). Nonetheless, the ongoing excessive, unauthorized, and often concealed uses of surveillance in government, the workplace, and the marketplace, seen especially from the late 20th century onwards continue to make the concept of surveillance politically contentious.

The above definition of surveillance may be used to understand the historical development of the term, its conventional and more controversial uses as a concept, and its ongoing critical capacities. Historically, 'surveillance' practices may be said to antedate the introduction of the concept of 'surveillance,' meaning that the concept may be applied to, for example, military intelligence, workplace supervision, and public 'policing'—also *avant la letter*—occurring from ancient times.

Increasingly, from the 19th century, it is the *technologies* used for surveillance that help to define the inherent changes in the modes of surveillance, that in turn require constant rethinking of the concept itself. Those technologies, themselves products of desires for improved communication, industrial production, or military prowess, became merged in the later 20th century, in 'information technology,' and latterly, on the internet, social media, and platform companies. Most recently, algorithmic analysis of extremely large datasets, artificial intelligence, and machine learning underlie many 'smart' surveillance activities, from fitness wearables to smart homes and cities (see Smart City by Hayward). Such issues are rapidly becoming crucial in surveillance debates (e.g., Crawford, 2021) including ones concerning both criminal and police use of AI.

This is why the concept of surveillance is not only required for but central to the digital context; the former has developed symbiotically with the latter. However, like its context, the digital, the practices of surveillance are means to other ends, rather than representing a human purpose in their own right. This may be demonstrated in each context where surveillance as social practices appear, which is why the practices are frequently controversial and the concept itself is contested. Here, the chosen window into the concept of surveillance is the burgeoning field of surveillance studies, within which criminology has played a substantial role from the outset.

The political-economic context and its accompanying technological features have always been significant aspects of whatever surveillance is practiced; to neglect them is to misunderstand both the phenomenon and the concept. Today, the digital

context, dependent on the internet and on complex algorithms, is central to surveillance. Data, in other words, is the means whereby human beings, in their many activities, are made visible, represented, and treated (Taylor, 2017). But although the concept of surveillance is rightly related to an infrastructural feature of contemporary societies, and is highly automated (Eubanks, 2017; Andrejevic, 2020) it also still refers to a set of *social practices* (Finn, 2012; Marx, 2016).

Related concepts

Several concepts are close neighbours of surveillance. One, 'spying,' is sometimes conflated with surveillance, unsurprisingly, due to the role of surveillance in intelligence gathering. The confusion is seen in former FBI Director James Comey denying the charge that the FBI spied on the Trump electoral campaign by placing it under electronic surveillance in 2019. "I have never thought of that as spying," he said (Kanno-Youngs and Schmidt, 2019). Unlike surveillance, one can argue that all spying involves secrecy, implying enmity or competition. A second concept is 'supervision,' which has similar roots as 'surveillance' but connotes not only observing, but also *directing the execution* of some activity or work. As we shall see, in a digital era surveillance is tending towards supervision in this sense, which means that further conceptual clarification is needed.

A third close concept is 'monitoring,' which also involves observation, often with the connotation of regular checking and reporting over time. In a workplace, for instance, employees may be monitored to check that their work is appropriate and satisfactory (Ball, 2010), but the workplace itself may also be monitored, for example, for health and safety or security. As Ball (2021: 11) observes, 'surveillance' and 'monitoring' may be used interchangeably in this context. However, the stress for those who use 'surveillance' is on power, politics, resistance, and meaning-making, whereas others are primarily concerned with the effectiveness—however defined—of monitoring.

If spying, supervision, and monitoring are close concepts to surveillance, then tracking and profiling should perhaps be added to the list. However, tracking and profiling, along with monitoring, are frequently used as concepts that specify what *aspects* of surveillance are under review. This is the case, for instance, in a recent book suggesting that 'tracking capitalism' might be a better term than Zuboff's 'surveillance capitalism' (Goldberg, 2021). Surveillance is in this sense an umbrella concept.

The concept of privacy is also associated with surveillance; sometimes it is seen as its antidote, if not its antonym (Stalder, 2002). Some engaged with regulating surveillance use 'privacy' as a key concept but may also quibble about using the concept of 'surveillance' in some contexts, such as marketing. Yet others argue that marketing both erodes autonomy and privacy *and* empowers consumers (Darmody and Zwick, 2020). Much debate hangs on how far privacy can cope with the *social*, as well as on individual aspects of privacy (Nissenbaum, 2009). Today, however, the digital environment frequently takes surveillance far beyond identifiable individuals and instead to-

ward the workings of a data infrastructure (Austin and Lie, 2021). The valuable concept of privacy thus becomes less germane to the full range of surveillance practices, reducing what was seen as its former larger congruence with the concept of surveillance. Following this, at the political level, pleas for privacy can only be a partial response to current surveillance practices.

The development of the 'surveillance' concept

The earliest meaning of the concept of surveillance, appropriate to its etymology, was that of *observation.* The 'watchman,' assigned to 'keep watch' in the city, was on duty in ancient times, until such watching was professionalized as a 'policing' task in 18th-century Europe. By 1829, Robert Peel established the Metropolitan Police in London, and interestingly, one of their roles was to *be visible* in 'preventive patrolling.' If watch-keeping was done in a military context, against an enemy, however, *concealment* was much more likely, as it would also be practiced in urban or national security contexts as 'secret policing.' And by the 20th century such secret policing became more frequently associated with covert government observation of populations in Russia after the 1917 revolution, or in Germany under the Third Reich. In this same period surveillance technologies including record-keeping were also adopted to enhance observational techniques (see Jeffreys-Jones, 2017; Lyon, 1994).

Equally, surveillance as observation also occurs in workplace settings, as it has, using different terms, for millennia. Employers' desire to check on the appropriate and timely fulfillment of work tasks is the purpose of surveillance. Here too, such observation became much more formalized with the development of industrial capitalism, especially with the expansion of factories, that typically entailed larger groups of employees under one roof. Direct observation by 'foremen' was gradually enhanced by technical means, prominently, to include information collected on workers (Beniger, 1989). Towards the end of the 19th century, not only the capitalist workplace, but also capitalism's marketplaces, also practiced surveillance, mainly by the collation of spending and preference information on consumers (Lauer, 2017; Igo, 2018) but also, now, through audio analysis (Turow, 2021). So, what began as the literal watching of bodies, in each sphere, has gradually morphed into the collection of data, thus permitting an 'image' of the person to be built by the surveillor.

The intervention of technology, then, enables a certain distancing, from observing bodies in space, to deducing aspects of their behavior, extrapolating future potentials, or enacting regulation from the information gathered about them. This process also enables a second sense of the concept of surveillance, the *sorting* of populations into categories of background and behavior, something that has become a key to conceptualizing surveillance (Lyon, 2003). Surveillance practices cluster people in social and spatial categories so that they can be represented and treated as members of such groups (see Categorization and Sorting by Franko). Foreign students in wealthy coun-

tries, for example, may be sorted and ranked by their 'desirability' as immigrants during the application process (Brunner, 2022).

The difference between observing and sorting may be elucidated by considering Foucault's (1975, 1977) famous description of the Panopticon prison, in which inmates are normalized into conformity with institutional expectations through constant 'inspection' by a watcher who is invisible to them. The covert aspect reappears in this version of the concept. Here, the success of surveillance hangs on the direct *observation* of bodies. However, earlier in his chapter on 'panopticism,' Foucault directed attention to 17th-century plagues, in which surveillance was carried out by the collection of information. Details of plague victims enabled control of the situation through *categorizing* them so that different groups were treated differently.

If the concept of surveillance has shifted from direct observation to include sorting, the increasing use of information technologies also facilitates a move away from concern with actual bodies to binary digits, or 'bits.' A third aspect of the concept is *digitized* surveillance. That is, the object of surveillance is less corporeal—the 'image' above—and more related to what is now called data. In the hands of Gilles Deleuze (1992), such a situation reduces further the association of surveillance with observed bodies, to one that refers merely to 'dividuals'; discrete bits of data rather than complex individuals. Rather than just being normalized, subjects of surveillance are pulled into the 'machine' of control, which is surveillance as management. As Haggerty and Ericson (2000) note, the body is as it were reconstituted—as consumer, employee, patient, and so on—to fit the surveillance 'assemblage,' which in itself is increasingly geared to predictive, future events.

Building on *digitized* surveillance is a fourth understanding of the concept—*datafied* surveillance or *dataveillance* (Clarke, 1988). This expansion of the concept of surveillance allows for the exploration of contemporary surveillance which in practice has become infrastructural for today's global societies. As van Dijck (2014) notes, dataveillance is 'continuous' as well as ubiquitous; it is always on, everywhere. Moreover, whereas earlier concepts of surveillance assumed that observation and sorting and even digitization began in distinct spheres, surveillance as dataveillance adds up to what van Dijck calls a "whole ecosystem of connective media" (p. 198).

This is expressed above all in the phenomenon of surveillance capitalism, where large companies monitor and profit from data produced by everyday activities online and in the physical world. Van Dijck's "ecosystem of connective media" is dominated by search engines such as Google, and social media platforms such as Facebook that use dataveillance as the basis of business, hence 'surveillance capitalism'—whether approached from political economy (Mosco, 2014; Foster and McChesney, 2014), computing (Clarke, 2019) or sociology and social psychology (Zuboff, 2019).

Note that the four senses of surveillance identified here are also *nested*—they refer to each other and each later one is dependent on the one that preceded it. Some kind of *observation* is required for categorizing and *sorting*; sorting is now *digitally* assisted, becoming part of the current *infrastructure* of surveillance.

Surveillance: a multidisciplinary concept

Because it is an inherently multidisciplinary concept, surveillance also has varying nuances of meaning within different disciplinary fields. Thus, for instance, its use in public health discourse and epidemiology is different from that in consumer marketing, and that in computing sciences from that of legal discourses in regulation and law. Even in the social sciences, such as sociology, psychology, political science, and cultural studies, the exact sense of the 'surveillance' concept may fluctuate. This calls for careful translation work as well as stimulating much-needed interdisciplinary debate.

The strictest use of the concept of surveillance, historically at least, is in the legal domain, where in the US it refers to "the act of observing another in order to gather evidence" which may be covert or overt (Legal Information Institute, 2021). This phrase situates surveillance in the realm of policing, although in this case 'surveillance' is prefixed with 'electronic.' In the European Union, the scope of surveillance is seen more broadly, assuming rather than adding the 'electronic' dimension. The Data Protection Supervisor (EU, 2021) notes that "technological progress in the past few decades have (sic) made monitoring, tracking and profiling practices easier, cheaper and more accurate." This reading of the concept includes, for instance, both the public sphere—such as security—and the private—such as targeted advertising. As with the social science-based understanding of the concept, then, the use of digital technologies inflects 'surveillance' significantly.

As noted earlier, the concept of surveillance is an umbrella sheltering a range of possible activities that often must be qualified for more precise use. Each of the four senses of the concept of surveillance mentioned above—observation, sorting, digitized surveillance, and dataveillance—reflects a technologically-enabled distancing from contact with actual human bodies, using cameras, telephones, computers, and other technologies. How this occurs, in different settings, also inflects the use of the concept in various disciplines. But changes in technology also spell a *return* to bodies, now understood as data-sources rather than as the objects of direct vision or audible signals, for instance through biometric technologies such as facial recognition or iris scans (see Facial Recognition by Fussey). As argued earlier, the dialectic movement between technology and surveillance now, at least partially, reunites the conceptual field.

This is especially true of the notion of surveillance capitalism that relates organizationally to platform companies in particular, and symbolically to the device of the smartphone. By turning the concept into a qualifier of 'capitalism,' the concept of surveillance undergoes another alteration as a societal or civilizational descriptor. In fact, discussion of surveillance capitalism offers further contemporary opportunities to rethink the concept of surveillance from several disciplinary perspectives. Disciplines such as political economy, sociology, computing sciences, geography, business studies, and others each have interests in how surveillance is parsed.

In the early 21st century several developments in particular warrant careful attention. One is the political economy of surveillance (Ball and Snider, 2019), referred to

above in the debates over 'surveillance capitalism,' especially in the form developed by Shoshana Zuboff (2019). Another is the rapid rise of postcolonial and decolonial theory (Breckenridge, 2014; Mbembe, 2003; McCoy, 2009), not least because many forms of surveillance that are apparent in the global north were first trialed in colonial regimes of the global south, but also because contemporary colonial situations depend heavily on surveillance (e.g., Zureik, 2016). Each of these is singularly significant to the concept of surveillance today, both in their own right, and also seen in relation to each other, as, for instance, varieties of surveillance capitalism proliferate in the so-called global south.

At least three further strands of surveillance research affect how the concept is construed: class, race, and gender. Discussions of surveillance capitalism cannot be severed from class relations (Foster and McChesney, 2014; Mosco, 2014; McQuade, 2018; Fuchs, 2012) and issues of colonialism are inseparable from those of racialization and surveillance (Benjamin, 2019; Browne, 2015). Artificial Intelligence and Machine Learning systems designed at the beginning of the pandemic, for example, were hotly debated by civil society and public health researchers, especially with regard to how and whether racial and ethnic data should be used to train modeling algorithms within COVID-19 prediction platforms (Singh, 2020; McKenzie, 2020; Choi et al., 2021). The deployment of AI in facial recognition systems, as another example, is fraught with racial biases, given, among other things, their propensity to misidentify women of color (Buolamwini and Gebru, 2018). As for gender, as well as a growing number of feminist studies of surveillance (Taylor, 2019; Dubrofsky and Magnet, 2015), questions of gender identity increasingly feature in surveillance studies (Ball et al., 2009; Abu-Laban, 2015; Kafer and Grinberg, 2019).

The spheres within which the concept of surveillance is used are diverse, for example in national security, policing, marketing, epidemiology, and public health. The concept may be controversial, for instance in marketing, but the practices and tools in that sphere so closely resemble surveillance in other areas, that using the term 'consumer surveillance' is warranted (Turow, 2021). Surveillance practices can even be denied in areas such as national security, especially after 9/11 and the Snowden revelations, when the NSA claimed that using 'metadata'—which is in fact very revealing—was not surveillant (Schneier, 2012; Lyon, 2015; Thompson and Lyon, 2021). Significantly, it is datafication and the internet which above all not only enable surveillance—as dataveillance—to occur on a mass scale, but also to exhibit similar features across different domains. Indeed, surveillance conducted by internet platforms produces data that is widely sought by government-related agencies (Srnicek, 2016).

Surveillance as a concept is often treated somewhat one-sidedly as having salience mainly for the activity of 'watching over,' by whatever means. Yet, especially today, when surveillance is no longer restricted to specific security or policing 'suspects' or 'targets,' but affects everyone, the *experience* of surveillance becomes an important feature of surveillance *effects.* Indeed, beyond this, the activities of those subject to surveillance in digital contexts increasingly make a difference to the surveillance itself. This occurs through a looping process (Hacking, 2006), in which surveillance subjects

become aware of being watched and may consequently change their behavior, thus making it all the more essential that this dimension be considered (Lyon, 2018). Thus, social psychology and cultural analysis also have insights for surveillance studies.

Lastly, recall that some of the most significant studies of surveillance occur within works of literature, film, and art. Moreover, these have in turn stimulated conceptual work in other fields. For instance, one of the earliest sociological studies of surveillance (Rule, 1974) is clearly influenced by George Orwell's classic novel, *Nineteen-Eighty-Four.* Of course, Orwell's Big Brother has inspired many other arts productions, including the TV series of the same name, which queries the experience of surveillance (McGrath, 2004). The most recent relevant novel at the time of writing is Dave Eggers' *The Every* (2021), a brilliant sequel to *The Circle* (Eggers, 2013). In art, surveillance is a seductive theme in many exhibitions, and it is a prominent muse in ZKM's *CTRL [SPACE]* (ZKM, 2001; Allen et al., 2010). Film, too, plays a major role in exploring the concept of surveillance; classics include *The Conversation* (1974) and *Minority Report* (2002), which serendipitously coincided with the post-9/11 understanding of predictive dataveillance (Kammerer, 2012). Later, the TV series *Black Mirror* (2011–2019) played a role in sharply alerting viewers to some negative dimensions of digital surveillance, and documentaries such as *Social Dilemma* (2020) expose aspects of surveillance capitalism.

Relevance and impact of the concept

The concept of surveillance has a multi-faceted relevance and impact. While acknowledging its early significance in the 19th century, its relevance is vastly greater today. The impact of computing developments in the mid-20th-century Cold War era considerably raised the profile of the concept of surveillance and the growth of the commercial internet in the 1990s elevated it further until it reached exponential levels with social media in the early 21st century. The attacks of 9/11 (Ball and Webster, 2003), the Snowden revelations (Lyon, 2015), the Facebook-Cambridge Analytica scandal (Bennett and Lyon, 2019), and above all the global COVID-19 pandemic (Lyon, 2022a), clearly illustrate this point. Each event stimulated explosive growth in surveillance, involving both government–corporate partnerships and ordinary citizen-consumers.

The social sorting dimensions of surveillance are crucial to each expansion, increasingly so as 'smart' data analysis is infrastructurally implicated. Social sorting occurs on large, medium, and small scales, from global corporations to police departments to micro-businesses. While certain efficiencies may thus be enhanced, such sorting also has a demonstrable tendency to create or exacerbate the vulnerability of some groups. This applies especially to low-income people, or those caught in the intersections between class-race-gender categories. The sorting dimensions of contemporary surveillance were noted early on by Oscar Gandy (2021) and elaborated upon subsequently by many others (e. g., Lyon, 2003, 2021).

The majority of surveillance activities today are data-dependent, and their outcomes are the product of data gathering, analysis, and use (Cheney-Lippold, 2017).

COVID-19 pandemic-driven technological developments illustrate this well. The hasty design and development of digital identity systems around the world is one such example. As governments worked closely with the private sector to develop smartphone-based identity and vaccine verification solutions, their rationale is at once a matter of mobility and of economic recovery. This speaks directly to the fact that modes of surveillance are frequently implicated in processes that affect the life chances and choices, and the conditions of freedom and fairness, of millions world-wide.

The smartphone is the primary device for surveillance activities today, built on the communications network of the internet, and enabling surveillance of a highly personal—identifiable—and geographically locatable kind. While this sprang from the identifying, tracking, and sorting of consumers, and was hugely enhanced by the advent of social media, the resultant data, and the methods of processing it, continue to leak into different spaces. Access to such data has been made possible to policing, security, administrative, and other agencies. Political responses demanded by the distinctive modes of surveillance emerging in the 21st century include basic rights relating to data-handling. Importantly, notions such as 'data justice' (Taylor, 2017; Dencik et al., 2019) and 'digital citizenship' (Isin and Ruppert, 2020) are gaining currency for their relevance to contemporary surveillance, alongside appeals for privacy and data protection (see Privacy and Data Protection by Bygrave).

Conclusion

Analytically, modifications to the concept of surveillance mentioned here are helping to confront new realities such as 'smart' and 'platform' surveillance. Innovative proposals, such as data justice, are also important because they inform policy and regulation, as well as public opinion, at a time when older policy concepts such as privacy and data protection (Puri, 2021) require careful overhaul (see, e. g., Lyon, 2022b). Future directions for the concept's usage would do well to follow the routes of recognizing the political economy of surveillance—seen in debates over surveillance capitalism—and the decolonial approaches that are illuminating not only the global south, but also in the global north, among once-colonizing nations. At the same time, each conceptual expansion contributes to the vital focus on growing vulnerabilities associated with current data-surveillance practices that are deepening inequalities of class, race, and gender (see Vulnerability by Ranchordas and Beck).

Surveillance is a contested concept, just because it is one of such great significance, especially in the present, and because alternative intellectual and political traditions view it differently. One seemingly intractable issue is whether the associations of surveillance with power and authority mean that its impacts are *inescapably* negative (Monahan, 2021; Harding, 2018; McQuade, 2018). Given the cognate evidence of how much surveillance continues to be dependent on military-security, rapacious capitalist, and white colonial forces, its dubious reputation seems well-deserved.

Those who take a different view argue that surveillance may be performed not only in benign fashion—such as in public health surveillance and even in some types of policing and security surveillance—but also positively, for the common good (Stoddart, 2021). The latter arguments depend, not on seeing surveillance through rose-coloured spectacles, but on recalling that the concept of surveillance always refers to social practices, and thus are subject to principled critique and open to political challenge. As criminologist Gary Marx has stated, "surveillance itself is neither good nor bad, but context and comportment make it so" (Marx, 2016).

Surveillance is also an inherently critical concept, one that alerts us to some of the most egregious injustices and entrenched power imbalances visible worldwide. But it is also increasingly complex and hidden, raising new challenges for empirical investigation. Critical researchers strive to make hidden surveillance data visible and legible to civil society. Equally, surveillance is critical because it questions the authority of those who argue in techno-solutionist (see Mozorov, 2014 for assessment) and technologically determinist (see Zuboff, 2015 for assessment) terms that dataveillance serves, primarily, the cause of human betterment.

Suggested reading

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance and Society*, 12(2), 185 – 196.

Lyon, D. (2018). *The Culture of Surveillance: Watching as a Way of Life.* Cambridge: Polity.

Monahan, T., & Murakami Wood, D. (2018). *Surveillance Studies. A Reader.* Oxford: Oxford University Press.

References

Abu-Laban, Y. (2015). Gendering surveillance studies: The empirical and normative promise of feminist methodology. *Surveillance & Society*, 13(1), 44 – 56. DOI: https:/doi.org/10.24908/ss.v13i1.5163

Allen, J., Robertson, K., & Smith, S. E. K. (eds.) (2010). *Sorting Daemons: Art, Surveillance Regimes and Social Control.* Kingston: Agnes Etherington Art Centre.

Andrejevic, M. (2020). *Automated Media.* London: Routledge.

Andrejevic, M., & Selwyn, N. (2022). *Facial Recognition.* Cambridge: Polity.

Austin, L., & Lie, D. (2021). Data trusts and the governance of smart environments: Lessons from the failure of Sidewalk Labs Urban Data Trust. *Surveillance & Society*, 19(2), 255 – 261. DOI: https:/doi.org/10.24908/ss.v14409

Ball, K. S. (2010). Workplace surveillance: An overview. *Labor History*, 51(1), 87 – 106. DOI: https:/doi.org/10.1080/00236561003654776

Ball, K. S. (2021). *Electronic Monitoring and Surveillance in the Workplace.* Report for the European Commission Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC125716

Ball, K., Phillips, D., Green, N., & Koskela, H. (eds.) (2009). Surveillance studies needs gender and sexuality. *Surveillance & Society*, 6(4) (special issue). DOI: https/doi.org/10.24908/ss.v6i4.3266

Ball, K., & Snider, L. (eds.) (2019). *The Surveillance-Industrial Complex: A Political Economy of Surveillance.* London: Routledge.

Ball, K., & Webster, F. (eds.) (2003). *The Intensification of Surveillance.* London: Pluto Press.

Beniger, J. (1989). *The Control Revolution: Technological and Economic Origins of the Information Society.* Cambridge, MA: Harvard University Press.

Benjamin, R. (2019). *Race after* Technology. Cambridge: Polity Press.

Bennett, C., & Lyon, D. (2019). Data-driven elections: Implications and challenges for democratic societies. *Internet Policy Review* 8(4). DOI: https/doi.org/10.14763/2019.4.1433

Breckenridge, K. (2014). *Biometric State: Global Politics of Identification and Surveillance in South Africa 1850 to the Present.* Cambridge: Cambridge University Press.

Browne, S. (2017). *Dark Matters: On the Surveillance of Blackness.* Durham NC: Duke University Press.

Brunner, L. R. (2022). Higher education institutions as eyes of the state: Canada's international student compliance regimes. *Globalisation, Societies and Education*, 20(1), 236 – 251. DOI: https/doi.org/10.1080/14767724.2022.2037407

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine-Learning Research* 81, 1 – 15.

Cheney-Lippold, J. (2017). *We are Data: Algorithms and the Making of our Digital Selves.* New York: New York University Press.

Choi, K. H., Denice, P., Haan, M., & Zajacova, A. (2021). Studying the social determinants of COVID-19 in a data vacuum. *Canadian Review of Sociology*, 58(2), 146 – 164. DOI: https/doi.org/10.1111/cars.12336

Clarke, R. (1988). Information technology and dataveillance. *Communication of the ACM*, 31(5), 498 – 512. DOI: https://doi.org/10.1145/42411.42413

Clarke, R. (2019). Risks inherent in the digital surveillance economy: A research agenda. *Journal of Information Technology*, 34(1), 59 – 80. DOI: https/doi.org/10.1177/0268396218815559

Crawford, K. (2021). *Atlas of AI.* New Haven: Yale University Press.

Darmody, A., & Zwick, D. (2020). Manipulate to empower: Hyper-relevance and the contradictions of marketing in the age of surveillance capitalism. *Big Data & Society*, 7(1).

Deleuze, G. (1992). A postscript on the societies of control. *October.*

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society*, 22(7), 873 – 881. DOI: https://doi.org/10.1080/1369118X.2019.1606268

Dubrofsky, R., & Magnet, S. (eds.) (2015). *Feminist Surveillance Studies.* Durham, NC and London: Duke University Press.

Eggers, D. (2013). *The Circle.* New York: Knopf/McSweeneys.

Eggers, D. (2021). *The Every.* New York: Vintage.

EU. (2021). Surveillance. https://edps.europa.eu/data-protection/our-work/subjects/surveillance_en

Eubanks, V. (2017). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York: St. Martin's Press.

Finn, J. (2012). Seeing surveillantly: Surveillance as social practice. In A. Doyle, R. Lippert, & D. Lyon (eds.), *Eyes Everywhere: The Global Growth of Camera Surveillance* (pp. 67 – 80). London: Routledge.

Foster, J. B., & McChesney, R. (2014). Surveillance capitalism: monopoly-finance capital, military-industrial complex and the digital age. *Monthly Review*, 66(3).

Foucault, M. (1975). *Surveiller et punir: Naissance de la prison.* Paris: Gallimard

Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison.* New York: Basic Books.

Fuchs, C. (2012). Political economy and surveillance theory. *Critical Sociology*, 39(5), 671 – 687. DOI: https://doi.org/10.1177/0896920511435710

Gandy, O. (2021). *The Panoptic Sort: A Political Economy of Personal Information.* New York: Oxford University Press. (Revised and expanded version of 1993 title)

Goldberg, D. T. (2021). *Dread: Facing Futureless Futures.* Cambridge: Polity Press.

Hacking, I. (2006). Making up people. *The London Review of Books*, 28(16).

Haggerty, K., & Ericson, R. (2000). The surveillance assemblage. *British Journal of Sociology*, 51(4), 605 – 622. DOI: https://doi.org/10.1080/00071310020015280

Harding, J. M. (2018). *Performance, Transparency and the Cultures of Surveillance.* Ann Arbor: University of Michigan Press.

Higgs, E. (2004). *The Information State in England: The Central Collection of Information on Citizens since 1500.* Basingstoke: Palgrave Macmillan.

Igo, S. (2019). *The Known Citizen: A History of Privacy in Modern America.* Cambridge, MA: Harvard University Press.

Isin, E., & Ruppert, E. (2020). *Being Digital Citizens* (2nd ed.). London: Rowman and Littlefield.

Jeffreys-Jones, R. (2017). *We Know All About You: The Story of Surveillance in Britain and America.* Oxford: Oxford University Press.

Kafer, G., & Grinberg, D. (2019). Editorial: Queer surveillance. *Big Data & Society*, 17(5), 592 – 601. DOI: https://doi.org/10.24908/ss.v17i5.13541

Kammerer, D. (2012). Surveillance in literature, film and television. In K. S. Ball, K. D. Haggerty, & D. Lyon (2012). *Routledge Handbook of Surveillance Studies* (pp. 99 – 106). London: Routledge.

Kanno-Youngs, Z., & Schmidt, M. S. (2019). Comey defends Trump campaign surveillance: 'I have never thought of that as spying'. *New York Times*, April 11, 2019. https://www.nytimes.com/2019/04/11/us/politics/comey-trump.html

Lauer, J. (2011). Surveillance history and the history of new media: An evidential paradigm. *New Media & Society*, 14(4), 566 – 582. DOI: https://doi.org/10.1177/1461444811420986

Lauer, J. (2017). *Creditworthy: A History of Consumer Surveillance and Financial Identity in America.* New York: Columbia University Press.

Legal Information Institute. (2021). Surveillance. https://www.law.cornell.edu/wex/surveillance

Lyon, D. (1994). *The Electronic Eye: The Rise of Surveillance Society.* Cambridge: Polity Press.

Lyon, D. (ed.) (2003). *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*, London: Routledge.

Lyon, D. (2007). *Surveillance Studies: An Overview.* Cambridge: Polity Press.

Lyon, D. (2015). *Surveillance After Snowden.* Cambridge: Polity Press.

Lyon, D. (2018). *The Culture of Surveillance.* Cambridge: Polity Press.

Lyon, D. (2021). Sorting. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.), *Uncertain Archives: Critical Keywords for Big Data* (pp. 477 – 484). Cambridge MA: MIT Press.

Lyon, D. (2022a). *Pandemic Surveillance.* Cambridge: Polity Press.

Lyon, D. (2022b). *Beyond Big Data Surveillance: Freedom and Fairness*, Research Report, Ottawa. https://www.sscqueens.org/sites/sscqueens.org/files/bds_report_eng-2022-05-17.pdf

Marx, G. T. (2016). *Windows into the Soul: Surveillance and Society in an Age of High Technology.* Chicago: University of Chicago Press.

Mbembe, A. (2003). Necropolitics. *Public Culture*, 15(1), 11 – 40.

McCoy, A. (2009). *Policing America's Empire: The United States, the Philippines and the Rise of the Surveillance State.* Madison: University of Wisconsin Press.

McGrath, J. (2004). *Loving Big Brother.* London: Routledge.

McKenzie, K. (2020). Race and ethnicity data collection during COVID-19 in Canada: If you are not counted you cannot count on the pandemic response. *The Royal Society of Canada,* November 12, 2020.

McQuade, B. (2018). *Windows into the Soul* or the clouded glass of surveillance studies? *Critical Sociology*, 44 (4 – 5), 815 – 824. DOI: https://doi.org/10.1177/0896920517751588

Monahan, T. (2021). Reckoning with COVID, racial violence, and the perilous pursuit of transparency. *Surveillance & Society*, 19(1), 1 – 10.

Mozorov, E. (2015). *To Save Everything, Click Here: The Folly of Technological Solutionism.* New York: Public Affairs.

Mosco, V. (2014). *To the Cloud: Big Data in a Turbulent World*, London: Routledge.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy and the Integrity of Social Life.* Stanford: Stanford University Press.

OED (*Oxford English Dictionary*). 2011. Surveillance. Oxford: Oxford University Press.

Orwell, G. (1949). *Nineteen Eighty-Four.* London: Secker and Warburg; New York: Harcourt, Brace.
Puri, A. (2021). A theory of group privacy. *The Cornell Journal of Public Policy*, 30(3), 477 – 538. https://heinonline.org/HOL/P?h=hein.journals/cjlpp30&i=501
Rule, J. (1974). *Private Lives and Public Surveillance: Social Control in the Computer Age.* New York: Shocken.
Schneier, B. (2012). *Liars and Outliers: Enabling the Trust that Society needs to thrive.* New York: Wiley.
Singh, S. (2020). Collecting race-based data during coronavirus pandemic may fuel dangerous prejudices. *The Conversation*, May 27, 2020.
Srnicek, N. (2016). *Platform Capitalism.* Cambridge: Polity.
Stalder, F. (2002). Privacy is not the antidote to surveillance. *Surveillance & Society*, 1(1), 120 – 124. DOI: https://doi.org/10.24908/ss.v1i1.3397
Stoddart, E. (2021). *The Common Gaze.* London: SPCK.
Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2). DOI:_https://doi.org/10.1177/2053951717736335
Taylor, L. (2019). Global data justice. *Communications of the ACM*, 62(6), 22 – 24.
Taylor, L. (2020). The price of certainty: How the politics of pandemic data demand an ethics of care. *Big Data & Society*, 7(2). DOI: https://doi.org/10.1177/2053951720942539
Thompson, S., & Lyon, D. (2021). Pixies, pop-up intelligence, and sandbox play: The New Analytic Model and national security research in Canada. In D. Lyon & D. Murakami Wood (eds.), *Big Data Surveillance and Security Intelligence* (pp. 112 – 125). Vancouver: University of British Columbia Press.
Tucker, H. (2017). *City of Light; City of Poison: Murder, Magic and the First Police Chief of Paris.* New York: W. W. Norton.
Turow, J. (2021). *The Voice Catchers: How Marketers Listen In to Exploit Your Feelings, Your Privacy and Your Wallet.* New Haven: Yale University Press.
Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197 – 208. DOI: https://doi.org/10.24908/ss.v12i2.4776
ZKM. (2001). *Rhetorics of Surveillance: From Bentham to Big Brother.* https://zkm.de/en/event/2001/10/ctrl-space-rhetorics-of-surveillance
Zuboff, S. (2015). Big Other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30, 75 – 89. DOI: https://doi.org/10.1057/jit.2015.5
Zuboff, S. (2019). *The Age of Surveillance Capitalism.* New York: Public Affairs.
Zureik, E. (2016). *Israel's Colonial Project in Palestine.* London: Routledge.

Katja de Vries

54 Synthetic data and generative machine learning

Abstract: Artificially generated informational outputs of generative ML models are called *synthetic.* Synthetic data is output that has analytic value for research or AI-model creation. Synthetic media is informational output that is consumed as content. Synthetic data are believed to be a promising solution to overcoming privacy, data scarcity, and bias concerns. The value of synthetic data in relation to these problems depends on the context. Synthetic media come in three categories: lawful, 'lawful but awful,' and illegal. Synthetic content can be regulated in legal or technical ways, or through human content verification.

Keywords: synthetic data, synthetic media, deepfakes, generative AI, regulation of synthetic outputs

Introduction

The generative Artificial Intelligence (AI) revolution in the late 2010s has been called the dawn of machine creativity and artificial imagination (see Artificial Intelligence by Van Brakel). While one might be critical of this anthropomorphic description, there is no denying that progress has been enormous. In the early 2020s generative machine learning (ML) models can do what was still unimaginable in the 2010s: produce new realistic and convincing variations on existing informational patterns generated by humans (conversations, newspaper articles, drawings, songs, grant applications, jokes, recipes, etc.) or real-world systems (human faces, satellite images, movement patterns, financial transactions, DNA sequences of living entities, etc.). Important developments in the field of generative ML, such as the invention of Variational Autoencoders (VAEs) (2013) and Generative Adversarial Networks (GANs) (2014), were the first steps in this revolution that allows models to successfully generalize beyond what they were trained on. In the early 2020s several large-scale generative models where created: these contain billions of parameters and require massive training data sets and enormous computational power (see Computation by Mazzilli Daechsel). In 2021 the term *foundation* model was coined to describe generative models "trained on broad data at scale" that are "adaptable to a wide range of downstream tasks" (Bommasani et al., 2021: 3). Next to some large-scale GANs and VAEs, two important types of foundation models are Large Language Models (LLMs) and Diffusion Models. LLMs, such as Open

Disclaimers and acknowledgments: This contribution is based on research conducted in the project "CreAI: Co-existing with creative Artificial Intelligence within the limits of EU law. Data protection, intellectual property, freedom of expression and cybercrime" (2020 – 2024) funded by the Ragnar Söderberg Foundation.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-055

AI's *Generative Pre-trained Transformer* (GPT) model or the *Language Model for Dialogue Applications* (LaMDA) powering Google's *Bard*, are best known for their text-generation. Diffusion Models, underlying applications such Stability AI's *Stable Diffusion* and Open AI's *DALL-E*, are well-known for their text-to-image generation.

Artificially generated informational outputs of generative ML models are called *synthetic.* This entry is devoted to two types of synthetic outputs: data and media. Synthetic data is output that mimics the statistical properties of collected data and thus has analytic value (Gal and Lynskey, 2023) for research or the training of AI models (Jordon et al., 2022). In contrast, synthetic media is informational output that is to be consumed as content.

Synthetic outputs: relevance for criminology and regulation

For the field of criminology synthetic outputs are important in several ways.

First, as a *tool.* Synthetic *data* can be of use in criminological research or as training examples for criminological AI models (predictive policing, lie detectors, etc.). Synthetic *media* can act as bait to capture criminals (such as 'Sweetie,' an avatar of a 10-year-old girl, used in 2013 to capture individuals paying online for pedophilic sexual acts). All such uses as a tool raise legal and ethical concerns (see below for more detail).

Secondly, synthetic outputs can be the *object of criminological study* in at least three ways. First, when a special subcategory of synthetic data that acts like an optical illusion for AI models (so-called *adversarial synthetic data*) is used to destabilize the formation of AI models or to confuse their application. This poses serious security risks and can result in various crimes. Second, when *synthetic media that mislead humans* are used for criminal purposes. Third, when some types of synthetic media, such as synthetic child pornography of non-existent children, operate in *moral grey zones* that require a rethinking of which behavior should be criminalized.

Depending on their use the same synthetic outputs can sometimes act as data and sometimes as media. Imagine a generative model that is trained on facial images of real convicts and generates synthetic varieties of non-existing ones. Let's call it '*Lombroso 2.0*'—as a wink to how such model resuscitates the long-rejected 19th-century theory of criminal facial traits. One could imagine several applications of this hypothetical model. Its outputs could be used as *data:* as a way to share sensitive data in a privacy-preserving way or to train a smart camera to recognize potential criminals based on facial characteristics. Its outputs could also be used as *media:* as entertaining or educational content for websites generating images of non-existing criminals (www.thiscriminaldoesnotexist.com) or merging faces with statistical traits generated by *Lombroso 2.0* (www.howwouldilookasacriminal.com).

In this entry I first discuss synthetic *data* as a tool for criminology, and as an *object of criminological study*. Then I discuss synthetic *media* as an *object of criminological study*. I conclude by discussing different modalities to *regulate* synthetic outputs.

Synthetic data as a tool for criminology

Sometimes respondents fail to answer a question in a survey, a sensor misses to record an input, etc. Many decades before the dawn of the AI revolution in the 2010s, statistics used *data imputation:* a technique to substitute missing data with statistical guesses about what the missing value could have been. Since the advances of generative AI in the late 2010s synthetic data are no longer merely placeholders for missing values but are considered to have the potential to solve three major challenges in data processing, research, and the creation of AI-models: privacy, scarcity of data, and bias.

First, *privacy:* the hope is that turning a real dataset into a synthetic one can be an effective way of anonymization (Bellovin et al., 2019). Normally anonymization is a process that entails that certain parts of the data have to be destroyed. This is called the privacy-utility trade-off: by removing information, data might become anonymized, but also less useful. When the use of collected data is infringing on privacy or data protection laws (see Privacy and Data Protection by Bygrave), synthetic data holds the promise of anonymous data that escapes the privacy–utility trade-off. For the creation of synthetic data, a generative AI model has to be trained on collected data. This model can then be used to generate synthetic (or fake) data with the same statistical properties. Once the training is completed the original dataset can be removed. To illustrate: in principle a data set with faces of non-existing convicted criminals generated by the aforementioned hypothetical model *Lombroso 2.0* would be considered as only containing non-personal data and fall outside the scope of any data protection requirements. However, there are critical voices (Stadler et al., 2022) about the capacity of synthetic data to escape the privacy–utility trade-off. In certain cases, the synthetic data might end up being too close to the original and the data would not meet the requirements of true anonymization. In other cases, the synthetic data might be so detached from the original data that they are no longer really representative or reliable. One well-known problem is mode collapse: a generative model generates too many data of one particular type—for example, *Lombroso 2.0* could end up making lots of variations on the same type of face.

Second, *data scarcity:* collecting data can be expensive, time-consuming, difficult, or sometimes impossible. Synthetic data holds the promise of feeding the multitude —a small data set can be transformed into a large one—when real data are not available with the required abundance. For example, in the training of autonomous vehicles the use of synthetic data is commonplace. Especially with regard to rare events such as car accidents there are simply not enough data available. Training on synthetic data can here be understood as learning from simulated events.

Third, *bias:* when data are not representative, databases can be complemented with synthetic data. For example, if it is likely that there is a bias in the number of arrests and convictions with regard to certain groups, this "gap between police records and 'true' level of crime" (Brunton-Smith et al., 2023: 2) could potentially be adjusted with synthetic data. While black males are underrepresented in most databases, they tend to be overrepresented in databases for law enforcement. In the case of the hypothetical *Lombroso 2.0* model this could mean that the presence of female and Caucasian faces can be boosted by adding synthetic varieties.

It is impossible to make a categorical claim about the value and reliability of synthetic data in resolving privacy, scarcity, and bias issues—it will largely depend on the specifics of a use context. However, in general it should be recognized that there are often risks involved with the creation of synthetic data that might limit their validity.

Synthetic data as an object of criminological study

Adversarial synthetic data (Goodfellow et al., 2017) are a type of optical illusion aimed at AI models: data is synthetically altered in a minimal way (imperceptible for the human eye) that leads AI to classify it in a completely incorrect way. Placing adversarial images in real life situations where classificatory AI models are at work can have disastrous outcomes: for example, a warzone with autonomous weapons that mistake a humanitarian aid camp for a terrorist shelter or a highway with self-driving cars that mistake a traffic sign for a cloud. Adversarial data cannot only undermine classificatory AI-systems during their use, but can also during the stage of model creation. If adversarial data are inserted into a training data set, the whole model can get completely unhinged. Poisoning training data with adversarial examples is not always illegal. An example of a legitimate use is *Nightshade*, a tool that can be used by artists to fight the use of their works as training material for image-generating AI models (Heikkilä, 2023). By letting *Nightshade* recreate their works into an adversarial data, artists can upload their works online while deterring from their use as training material for image-generating AI models.

Synthetic media as an object of criminological study

Synthetic media is AI-generated informational output that is consumed as content. From a criminological perspective an important subcategory within synthetic media is *deepfakes* (Chesney and Citron, 2019; Van der Sloot and Wagensveld, 2022). Article 3(60) of the European AI Act 2024/1689 defines deepfakes as "AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful." Thus,

according to the AI Act *deepfakes* always refer to someone or something that really exists.

One can distinguish three types of synthetic media: lawful, illegal, and 'awful but lawful'.

Lawful media. An example of lawful synthetic media is the synthetic chatbot alter-ego ('AI Yoon') of Yoon Suk-yeol, who became president of South-Korea in May 2022. Yoon used his synthetic alter-ego in his presidential campaign to increase his appeal with younger voters. In contrast to his public image of a stern, middle-aged conservative politician, Yoon's synthetic AI double seemed more human: using memes, jokes, and satire that went quickly viral.

Illegal media. Within many jurisdictions with a constitutional protection of freedom of expression, for example the EU and US, there is only a limited type of illegal expressions. This covers synthetic and natural speech alike: notable examples are incitement to hatred (see Hate Crime and Networked Hate by Powell, Stratton, and Cameron), the dissemination of racist statements, hate speech, holocaust denial, defamation (libel and slander), identity fraud, extortion, unlawful privacy breaches, and child pornography. An example of an illegal *deepfake* is the use of a synthetic voice for criminal impersonation, for example in kidnapping, accident, or banking scams where individuals receive a phone call and are asked to transfer money. Many political or sexual deepfakes such as revenge porn, porn with the faces of celebrities added, or deepfakes that make politicians say things that damage their image, can be treated under existing criminal law. Within synthetic pornography one can distinguish between content that stars identifiable people and content that is unrelated to any such individuals. In jurisdictions where pornography is legal, most of the material with identifiable people is likely to be illegal if they did not consent to their appearance, particularly if the material harms a person's reputation (defamation, privacy, or data protection). In contrast, most explicit sexual material that contains non-existing people and that cannot be related to an identifiable person, would in principle be considered lawful in jurisdictions where pornography is permitted, unless there is an obscenity law that prevents a particular expression.

'Awful but lawful' media. There is sexually explicit material that many experience as obscene or revolting but that is not illegal. There is disinformation (intentionally misleading) or misinformation (not intentionally misleading) that is experienced by many as antidemocratic or offensive but that is not illegal. Most of the 'law but awful' content is something that simply has to be tolerated by a democratic society. Only disinformation is subject to some regulation, albeit in a soft manner—for example through support of media literacy and a voluntary self-regulatory code of practice for relevant industry players (European Commission 2022), which is given some legal bite as it is seen as an indicator of compliance under the EU Digital Services Act 2022/2065. However, the pushdown on disinformation should be cautious, and not be confused with the fight against illegal content.

At the boundary of the illegal and lawful-but-awful category are some types of synthetic media where the legislator might need to reconsider legality. Some child sexual

abuse material (CSAM) creates such moral grey-zone questions and tensions with freedom of expression. Recently there has been a boom in synthetic CSAM, both starring identifiable real children and material unrelated to any existing child. CSAM is considered illegal in most jurisdictions in the world. When CSAM is created with real children the harm is obvious—the child is harmed. However, what if no real children are involved, such as is the case with artificial CSAM (animations, manga, etc.) and pretend CSAM (where actors are adults pretending to be children)? Such material could be considered harmful following the 'whetting the appetite' argument—watching CSAM could potentially inspire sexual actions towards real children. There is no conclusive scientific evidence that watching CSAM actually translates to an increase in sexual child abuse (Bernstein, 2023) and this means different jurisdictions have different positions as to whether artificial and pretend content fall under the definition of CSAM. For example, in Article 20(3) of the Lanzarote Convention (*Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse*, CETS N°201, 25 October 2007) signatory parties are given the option to exclude artificial and/or pretend content. An additional nuance that is made in some jurisdictions is that within the category artificial CSAM a difference is made between realistic non-existent children and those that are clearly figments of imagination. The definition of 'child pornography' in Article 2(c) of the Child Pornography Directive 2011/93/EU has been interpreted to cover the former, but not the latter. One reason to criminalize synthetic CSAM of realistic non-existent children is that law enforcement wastes time and energy on trying to identify them.

In 2023 a District Court in Sweden (Tingsrätt Skaraborg, B 1428–23, 21 July 2023) found a man guilty of an *attempt* to create synthetic CSAM. Even though the imagery was not realistic (for example, a naked child with three legs) the Court deemed that the prompts showed an *intent* to generate realistic CSAM.

Different modalities to regulate synthetic outputs

Three main approaches can be distinguished in regulating synthetic output: pre-emptive watermarking, use of AI-fuelled detection models, and content-based approaches.

The first approach relies on making transparent that content is AI-generated. For example, in China the *Provisions on the Administration of Deep Synthesis Internet Information Services* (11 December 2022) state that all providers of systems that generate AI-output should ensure that it is be watermarked. Article 50(4) of the EU AI Act 2024/1689 mandates that deployers of systems "shall disclose that the content has been artificially generated or manipulated." The provisions build on a similar sentiment, though the scope of the Chinese provision is wider: Article 50(4) only applies to deepfakes, that is, to AI-generated material that pretends to be real and is misleading about its AI-generated nature.

The second approach is to train AI models to distinguish synthetic from collected material. There are different ways to do this. First, one can follow the spam-filter ap-

proach and train AI on existing (collected or synthetic) examples of deepfakes, or more broadly, synthetic material. The problem here is that synthetic material is changing all the time and that novel, so-called 'zero-day,' synthetic material goes undetected. One way to approach this issue is to compare text and image: for example, if a text mentions an existing person the model could check if the image corresponds to real imagery of that person (Reiss et al., 2023). For certain types of AI-generated content, such as texts, the metrics 'perplexity' and 'burstiness' can be used to establish how likely it is that a text is of synthetic nature. A low perplexity score means that a language model can easily predict the next word. A low burstiness score means that sentences are uniform in structure and length, and that there are no abrupt bursts of stylistic change. Low scores can thus indicate that a text is synthetic—or is human-written in a predictable way.

The last approach is based on human content verification and debunking, which can include decentralized collaborative verification by professional journalists, engaged citizens, or other trusted flaggers. While being significantly more labor-intensive than automated synthetic content filters, the benefit of human content verification is that it is not solely focused on if the material is of synthetic or human origin, because in the end that is not what separates legal from illegal content (de Vries, 2020; Jacobsen and Simpson, 2023).

Main takeaways

- Artificially generated informational outputs of generative ML models are called *synthetic.* Synthetic data is output that has analytic value for research or AI-model creation. Synthetic media is informational output that is consumed as content.
- Synthetic data are believed to be a promising solution to overcoming privacy, data scarcity and bias concerns. The value of synthetic data in relation to these problems depends on the context.
- Adversarial synthetic data are optical illusions that can unhinge AI models and pose security risks.
- Synthetic voices are increasingly used in illegal scams.
- Most deepfakes can be treated under existing criminal law.
- Synthetic media come in three categories: lawful, 'lawful but awful,' and illegal. The first two categories fall under the protective scope of freedom of expression.
- Synthetic child sexual abuse material with non-existing children is an example of synthetic media operating in moral grey zones that requires a rethinking of what behavior should be criminalized.
- Synthetic content can be regulated in legal or technical ways (pre-emptive watermarking, automated synthetic content detection, etc.), or through human content verification. The latter has the benefit that the focus is distinguishing legal from illegal content, instead of fetishizing the synthetic-real distinction too much.

Suggested reading

de Vries, K. (2020). You never fake alone: Creative AI in action. *Information, Communication & Society*, 23(14), 2110 – 2127. https://doi.org/10.1080/1369118X.2020.1754877

Gal, M., & Lynskey, O. (2023). Synthetic data: Legal implications of the data-generation revolution. 109 *Iowa Law Review*, Forthcoming, LSE Legal Studies Working Paper No. 6/2023. Available at SSRN: https://ssrn.com/abstract=4414385 or http://dx.doi.org/10.2139/ssrn.4414385

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). *Synthetic Data—what, why and how?* Report commissioned by the Royal Society and the Alan Turing Institute. *arXiv:2205.03257.*

Van der Sloot, B., & Wagensveld, Y. (2022). Deepfakes: Regulatory challenges for the synthetic society. *Computer Law & Security Review*, *46*, 1 – 15.

References

Legal sources

Regulation (EU) 2024/1689 of the European Parliament and the Council on 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

Regulation (EU) 2022/2065 of the European Parliament and the Council on 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

European Commission. (2022). *The Strengthened Code of Practice on Disinformation.* Brussels.

Other sources

Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258.*

Bellovin, S. M., Dutta, P. K., & Reitinger, N. (2019). Privacy and synthetic datasets. *22 Stanford Technology Law Review*, 1, 1 – 51.

Bernstein, D. (2023). Could AI-generated porn help protect children? *Wired.* https://www.wired.com/story/artificial-intelligence-csam-pedophilia/

Brunton-Smith, I., Buil-Gil, D., Pina-Sánchez, J., Cernat, A., & Moretti, A. (2023). Using synthetic crime data to understand patterns of police under-counting at the local level'. *CrimRxiv.*

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107, 1753 – 1819.

de Vries, K. (2020). You never fake alone: Creative AI in action. *Information, Communication & Society*, 23(14), 2110 – 2127. https://doi.org/10.1080/1369118X.2020.1754877

Gal, M., & Lynskey, O. (2023). Synthetic data: Legal implications of the data-generation revolution. 109 *Iowa Law Review*, Forthcoming, LSE Legal Studies Working Paper No. 6/2023. Available at SSRN: https://ssrn.com/abstract=4414385 or http://dx.doi.org/10.2139/ssrn.4414385

Goodfellow, I., Papernot, N., Huang, S., Duan, Y., Abbeel, P., & Clark, J. (2017). Attacking machine learning with adversarial examples. *OpenAI blog.* https://blog.openai.com/adversarial-exampleresearch/

Heikkilä, M. (2023). This new data poisoning tool lets artists fight back against generative AI. *MIT Technology Review.*

Jacobsen, B. N., & Simpson, J. (2023). The tensions of deepfakes. *Information, Communication & Society*, 27(6), 1095 – 1109. DOI: 10.1080/1369118X.2023.2234980

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). *Synthetic Data—what, why and how?* Report commissioned by the Royal Society and the Alan Turing Institute. *arXiv:2205.03257.*

Reiss, T., Cavia, B., & Hoshen, Y. (2023). Detecting deepfakes without seeing any. *arXiv:2311.01458.*

Stadler, T., Oprisanu, B., & Troncoso, C. (2022). Synthetic data – anonymisation Groundhog Day. In *31st USENIX Security Symposium (USENIX Security 22)*, 1451 – 1468.

Van der Sloot, B., & Wagensveld, Y. (2022). Deepfakes: Regulatory challenges for the synthetic society. *Computer Law & Security Review*, *46*, 1 – 15.

Dana Wilson-Kovacs

55 Translation

Abstract: This chapter focuses on the concept of *translation* and its development in sociology and science and technologies studies, illustrating how it can help criminologists analyze different forms of digitized control. The chapter explains how *translation* can be used as a methodological device and conceptual framework to illuminate forensic practices in police investigations, and more generally the adoption and use of technologies in policing. Translational approaches can provide a multi-perspectival understanding of the contexts in which digitized control takes place, the technologies that help monitor it, and the justifications, motives, and practices that accompany the development and implementation of such technologies.

Keywords: translation, technologies, forensic support, police investigation, forensic evidence

Policing is increasingly dependent on digital technologies: from body-worn cameras to forensics databases and the extraction of digital evidence from mobile devices, such technologies contribute in different ways to criminal justice outcomes. Their utility, efficiency, reliability, and overall benefits need to be explained to various stakeholders, and their operation sometimes adapted for different law-enforcement scenarios. Often social science scholars refer to these processes as 'translation.' Originally theorized in the linguistic field in the 1950s (Brisset, 2010), translation has been mainly understood as transferring meaning from one language to another. However, the concept of 'translation' has also been used to describe specific processes in domains such as medicine and health care (e. g., Straus et al., 2013), anthropology (e. g., Asad, 1986), organizational theory (e. g., Wæraas and Nielsen, 2016), sociology, science, and technologies studies (STS—e. g., Woolgar et al., 2009), and criminology (e. g., Grieco et al., 2014; Pesta et al., 2019). Here, to translate something refers to "a complex process of negotiation during which meanings, claims and interests change and gain ground" (Wæraas and Nielsen, 2016: 237).

In criminology, meanings attributed to the term vary: in Anglo-American scholarship, translation typically refers to collaborative enterprises where academic researchers and practitioners disseminate research findings to other criminal justice stakeholders (Bales et al., 2014), although it has been used in other contexts as well, for instance to provide a conceptual framework for the explanation of penal practices (Carrabine,

Corresponding author: m.d.wilson-kovacs@exeter.ac.uk, University of Exeter, Department of Social and Political Studies, Philosophy and Anthropology.

Disclaimers and acknowledgments: The support of the British Academy and of the Economic and Social Research Council (Research Grant ES/R00742X/1) are gratefully acknowledged.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-056

2000). Translational criminology seeks to demonstrate how academic research is used in evidence-based criminal justice policies and practices and can influence decision-making processes (Grieco et al., 2014). Topical literature here examines how researcher–practitioner partnerships may lead to better criminal justice outcomes, how law enforcement agencies use research to guide their decision-making and/or the barriers that prevent the use of such research (e.g., Telep and Winegar, 2015; Pesta et al., 2019).

While this preoccupation with the users of academic research is longstanding and befitting to criminological research (Holdaway and Rock, 1998), here we discuss translation as an extension beyond this established meaning to assist with the analysis of diverse forms of digitized control. In this sense, translation can help illuminate the contexts in which such control takes place, the technologies that help monitor it, and the motives and justifications that accompany the development and implementation of such technologies. Translation is typically multi-perspectival and accounts for the complex networks of exchanges and dependencies between the actors involved. A translational approach builds on insights from sociology and STS to study crime in digitized societies and further assist the development of criminology as a discipline (Heidensohn, 1998).

Understanding translation from a sociological lens

In sociology and STS, translation was originally articulated as part of Actor-Network-Theory (ANT), developed by French sociologists Michel Callon (1980, 1984) and Bruno Latour (1987), who used this framework to explain how actors seek to convince others to adopt their point of view in contexts such as technological innovations and scientific controversies, where there are divergent meanings and conflicting interests. Although Callon's and Latour's views on translation ultimately diverge, they shared an understanding of it as illustrating how power relations are constructed and maintained, in terms of "networks, alliances, points of resistance, instability and relative durability" (Carrabine, 2000: 313). Translation involves many different actors and is a complex process of negotiation "during which meanings, claims and interests change and gain ground" (Latour 1984: 237). The actors' own interests can be redefined and translated into networks of associations that result in the stabilization of scientific and technical objects. As such, establishing the latter does not depend on the usefulness of such objects, but on the successful building and maintenance of networks, structures, and associations that hold them together.

Callon's well-known example focused on the domestication of scallops and the fishermen of St Brieuc Bay (1984), where he identified four stages of the translation process: *problematisation*—where actors seek to convince others that they have the correct solutions and can establish obligatory passage points through them, *interressement*—where these actors "lock the other actors into the role that has been proposed to them" (1984: 190), *enrolment*—"a set of strategies in which researchers seek to define and interrelate the various roles they had allocated to others" (1984: 190) and *mobili-*

sation—"a set of methods used … to ensure that supposed spokesmen for various relevant collectivities are able to properly represent those collectivities" (1984: 190). Consider, for instance the development and implementation of software applications for predictive policing (see Prediction by Ķīlis, Gundhus, and Galis) The *problematisation* stage is illustrated by identifying the need to respond to incidents faster and more efficiently with the use of new technologies. The *interresement* stage typically articulates the values such technological applications can bring to policing, while *enrolment* seeks to gain the buy-in from those tasked with commissioning such technologies, usually policing users at senior level. Finally, *mobilisation* helps align the interests of law enforcement agencies with those of manufacturers and various other stakeholders, in order to bring the adoption of software applications into everyday policing.

Translation implies "transformation, deformations and dislocations" (Carrabine, 2000: 312). It involves both human and non-human actors and produces "convergences and homologies by relating things that were previously different" (Callon, 1980: 211). This understanding of translation has geometric, political, and semiotic meanings: geometric in the sense that different actors are mobilized in different directions for multiple ends, political in the sense that such actors will pursue their interests through power play, persuasion, and strategic manoeuvring, and semiotic because meanings transform as objects move across and between networks (Wæraas and Nielsen, 2016).

Stabilization is dependent on the ability to coordinate and manage how multiple and divergent actors with conflicting agendas reach consensus, even if this coordination is transient. Concepts such as 'boundary objects' (Star and Griesemer, 1989) and 'standardized packages' (Fujimura, 1992) have been proposed to describe "the interfaces between multiple social worlds" that enable "the flow of resources (concepts, skills, materials, techniques, instruments) among multiple lines of work" (Fujimura 1992: 170). Successful translations are accomplished through this tacit, invisible work of articulation, that involves "assembling, scheduling, monitoring, and coordinating all of the steps necessary to complete a production task" (Gerson and Star, 1986: 265). From laboratory studies (e. g., Knorr Cetina, 1999) to technologies (e. g., Pinch and Bijker, 1984), translation has helped unravel different epistemic cultures, their interactions and overlaps, and their taken-for-granted objects and practices. It has brought to fore the multiplicity of players, the strategies they follow and the alliances they mobilize, highlighting the negotiated, interactional nature of consensus formation and knowledge production. Applied to the realm of digital criminology, a translational approach can help criminologists explore in greater detail the situatedness and embeddedness of black-boxed (digital) technologies and practices and the ways in which different epistemic cultures and communities of practice contribute to their stabilization.

From employing AI in predictive policing, to the scrutiny afforded by automatic face recognition software, the monitoring provided by CCTV, and the identification of victims and suspects using digital and forensic genetic technologies, the technical has opened new areas of topical concern to criminological scrutiny. As sociological and STS scholarship has long established, the technical is never neutral and translation as a methodological approach can help capture the complexity and hybridity of its en-

tanglement. Egbert and Leese, for instance, build on Callon's and Latour's notion of translation to examine the 'hinges' (2021: 45) of predictive policing, more specifically the ways in which crime becomes digitally represented by data and algorithmically produced risk is established, communicated, and acted upon by patrol officers. Their processual understanding of predictive policing uses translation to highlight the gaps between domains and to outline how different logics become aligned. It demonstrates how an analytical approach focused on chains of translation (Latour, 1999) can aptly analyze how predictive police practices are established and knowledge and power are produced by alignments between data, software, analysts, briefings, maps, and patrol cars (Egbert and Leese 2021: 61). Here, configuring the sites and modes of translation appropriately is instrumental to its success. For Egbert and Leese (2021), officers, tasks, technologies, institutions, organizational discourses, and everyday practices must align so their rationales and operational capacities can produce what is known as predictive policing.

Translation in the social study of forensics and its use in policing

Using a broadly defined translational approach in the social science study of forensics can bring to fore the complexities behind the construction of forensic technologies and the recognition of expertise. Cole's work (2001), for instance, highlights the complex negotiations surrounding the acceptance, legitimacy, and judicial recognition of fingerprinting as a legitimate tool of identification. Similarly, the acceptance of DNA profiling in the US depended on overcoming the lack of early scientific consensus. DNA legitimacy was established via the standardization of its analysis (Derksen, 2000), its court acceptance (Lynch et al., 2008) and making the "FBI and its Hae III technology the obligatory passage point through which all actors had to pass if they wished to proffer DNA evidence in the American legal system" (Aronson, 2008: 201). Such translation extends to the implementation of new technologies, for instance in the deliberations surrounding the introduction of Rapid DNA solutions in policing in England and Wales (Wilson-Kovacs, 2022). While the need for the adoption of this relatively new forensic genetic tool for crime investigation has been articulated, justified, and resisted by various stakeholders, its introduction will also depend on its successful alignment to new public management demands for accountability, efficiency, and value for money.

At a macro-level, using a translational approach helps elucidate the motives, interests, and actions accompanying the development and use of technologies. At a micro-level, concepts such as 'alignment' and 'articulation' help explain how consensus is achieved between different actors and meanings are stabilized (e.g., Kruse 2020, 2021). This is particularly important in the case of forensic evidence and its journey through different epistemic cultures. The crime scene is a case in point: a space where actors from different epistemic cultures meet literally (e.g., police officers,

crime scene examiners, and publics) and symbolically (e.g. forensic scientists, jurors, and counsel). Kruse's analyses (2020, 2021) demonstrate how in their work, Swedish crime scene examiners (CSEs) translate their findings as well as mediate between the different epistemic cultures, by undertaking invisible articulation work to reconcile different viewpoints and stabilize the foundation upon which forensic evidence is to be built. Their simplified, standardized ways of presenting their work limits the frictions between the many actors involved.

Likewise, Wyatt (2014) discusses how in the investigation of volume crime in England and Wales, CSEs render the complexity of the crime scene into an ordered and coherent narrative for the officers involved in the investigation. This often produces the only institutionally sanctioned version of the crime scene and encapsulates many of the processes involved—from defining the limits of the crime scene itself, to assessing the risks of entering it, to complex decision-making about what may or may not be relevant to the investigation. In a similar way, Bechky (2021) shows how US criminalists working in forensic biology, chemistry, toxicology, and comparative evidence laboratories, translate their findings for lay and legal audiences (see Labs by Mazzilli Daechsel). Here, a broader understanding of translation is present, with CSEs and criminalists as the conduit between separate epistemic cultures.

Translation in digital forensics (DF)

While much of the social science scholarship on forensic evidence has focused on DNA and its social and ethical challenges (see DNA/Big Genome Data by Kaufmann), research that accounts for the production and use of digital forensic (DF) evidence is emerging. The DF community is an obligatory passage point in the creation of digital evidence, whose production resides at the intersection of competing commercial, governmental, and occupational interests. Divisions of labor, hierarchies of expertise, and inter and intra-professional exchanges are key to this production and can aid or disrupt the transformation of digital information into evidence (Wilson-Kovacs, 2021).

In an increasingly complex and rapidly changing technological landscape, with new types of offenses emerging, translation helps decipher the intricate arrangements, activities, and exchanges that assist evidential processes. Rappert et al.'s (2023) analysis of how police process indecent images of children offenses, for instance, discusses DF as a socially negotiated, collectively produced, and institutionally maintained set of practices. Documenting the formal and informal mechanisms through which large sets of data become discrete pieces of evidence, they question existing linear understandings of DF data processing and highlight the inherent difficulties of translating digital trace into usable information.

Similar to how chains of translation allow knowledge and power to travel through the organization from the desk of the analyst to the street level (Egbert and Leese 2021: 209), DF evidence is the result of sociotechnical practices that facilitate the successful transformation of digital information into evidential or intelligence data (see Intelli-

gence by Gundhus and Lundgaard). Its production depends on the standardized laboratory routines, guidelines, protocols, and standard operating procedures employed, and the documentation of the movement of evidential exhibits.

Analyses of the latter in the context of physical evidence show how they mask numerous actions involved in the flow of information: a photocopied signature on the form used to record the chain of custody, and the movements of samples between a police force and a laboratory before the completion of the chain, can bring to question the integrity of the sample and the admissibility of a DNA match obtained from its analysis (Lynch et al., 2008). Likewise, errors (see Error by Aradau), omissions and gaps in the officers' submissions of digital devices to the DF laboratory for analysis raise similar concerns. Unlike CSEs who typically rely on their knowledge and expertise to record a crime scene, DF examiners depend on the information provided by officers to complete their technical work. Wilson-Kovacs and Wilcox (2022) show how in DF officers often fail to include basic parameters to their requests for the examination of seized digital devices of suspects, witnesses, and victims. Translating the needs of officers into achievable investigative strategies and concrete results requires careful alignment of the parameters of a case with the most suitable method to produce a tailored analysis. It also involves communicating the results of the analysis back to the investigating officers, explaining their relevance and their methodological limitations. This more literal labour of translation and alignment illustrates the central role these actors play.

Conclusion

Translation can be broadly defined as the process through which actors construct common meanings and engage in continuous negotiation to achieve individual and collective objectives. Methodologically, a translation approach in digital criminology can help map the communities of practice involved and their multiplicity of interests, establish how technologies and information are stabilized, and explain the visions and expectations accompanying the design, development, and adoption of various digital tools in the prevention and investigation of crime. In a context where the production of evidence is increasingly black-boxed and made invisible by the technologies that assist with its creation, thinking through a translation lens about the trajectory of digital evidence from crime scene to court helps unpack the ways in which data are conceived, measured, and employed (Kitchin and Lauriault, 2014).

From the design, development, and adoption of digital tools in policing, to understanding how courts make use of new technological possibilities, translation can enhance empirical examinations of processes of evidence production and assist with the analysis of transnational co-operations between multiple national law enforcement agencies, as well as it can be equally beneficial to unpacking other facets of crime and justice in the face of ever-growing digitization.

Suggested reading

Bechky, B. A. (2021). *Blood, Powder, and Residue: How Crime Labs Translate Evidence into Proof.* Princeton, NJ: Princeton University Press.

Kruse, C. (2020). Swedish crime scene technicians: Facilitations, epistemic frictions and professionalization from the outside. *Nordic Journal of Criminology*, 21(1), 67 – 83.

References

Aronson, J. D. (2008). Creating the network and the actors: The FBI's role in the standardization of forensic DNA profiling. *BioSocieties*, 3(2), 195 – 215.

Asad, T. (1986). The concept of cultural translation in British social anthropology. In J. Clifford & G. Marcus (eds.), *Writing Culture: The Poetics and Politics of Ethnography* (pp. 141 – 164). Berkeley: University of California Press.

Bales, W. D., Scaggs, S. J., Clark, C. L., Ensley, D., & Coltharp, P. (2014). Researcher–practitioner partnerships: A case of the development of a long-term collaborative project between a university and a criminal justice agency. *Criminal Justice Studies*, 27(3), 294 – 307.

Bechky, B. A. (2021). *Blood, Powder, and Residue: How Crime Labs Translate Evidence into Proof.* Princeton, NJ: Princeton University Press.

Brisset, A. (2010). Cultural perspectives on translation. *International Social Science Journal*, 61(199), 69 – 81.

Callon, M. 1980. Struggles and negotiations to define what is problematic and what is not: The socio-logic of translation. In K. D. Knorr, R. Krohn, & R. Whitley (eds.), *The Social Process of Scientific Investigation* (pp. 197 – 219). Dordrecht: Springer Netherlands.

Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review*, 32(1 suppl), 196 – 233.

Carrabine, E. (2000). Discourse, governmentality and translation: Towards a social theory of imprisonment. *Theoretical Criminology*, 4(3), 309 – 331.

Cole, S. A. (2001). *Suspect Identities: A History of Fingerprinting and Criminal Identification.* Cambridge, MA and London: Harvard University Press.

Derksen, L. (2000). Towards a sociology of measurement: The meaning of measurement error in the case of DNA profiling. *Social Studies of Science*, 30(6), 803 – 845.

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London: Routledge.

Fujimura, J. H. (1992). Crafting science: Standardized packages, boundary objects, and "translation". In A. Pickering (ed.), *Science as Practice and Culture* (pp. 168 – 211). Chicago: University of Chicago Press.

Gerson, E. M. and Star, S. L. (1986). Analyzing due process in the workplace. *ACM Transactions on Information Systems (TOIS)*, 4(3), 257 – 270.

Grieco, J., Vovak, H., & Lum, C. (2014). Examining research–practice partnerships in policing evaluations. *Policing: A Journal of Policy and Practice*, 8(4), 368 – 378.

Heidensohn, F. (1998). Translations and refutations: An analysis of changing perspectives in criminology. In S. Holdaway & P. Rock (eds.), *Thinking about Criminology* (pp. 40 – 54). London: Routledge.

Holdaway, S., & Rock, P. E. (eds.) (1998). *Thinking about Criminology.* London: Routledge.

Kitchin, R., & Lauriault, T. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. In J. Thatcher, J. Eckert, & A. Shears (eds.), *Thinking Big Data in Geography* (pp. 3 – 20). Lincoln, NE: University of Nebraska Press.

Knorr Cetina, K. (1999). *Epistemic Cultures: How the Sciences make Knowledge.* Cambridge, MA: Harvard University Press.

Kruse, C. (2020). Swedish crime scene technicians: Facilitations, epistemic frictions and professionalization from the outside. *Nordic Journal of Criminology*, 21(1), 67 – 83.

Kruse, C. (2021). Attaining the stable movement of knowledge objects through the Swedish criminal justice system: Thinking with infrastructure. *Science & Technology Studies*, 34(1), 2 – 18.

Latour, B. (1984). The powers of association. *The Sociological Review*, 32(1_suppl), 264 – 280.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society.* Cambridge, MA: Harvard University Press.

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies.* Cambridge, MA: Harvard University Press.

Lynch, M., Cole, S. A., McNally, R., & Jordan, K. (2008). *Truth Machine: The Contentious History of DNA Fingerprinting.* Chicago: University of Chicago Press.

Pesta, G. B., Blomberg, T. G., Ramos, J., & Ranson, J. A. (2019). Translational criminology: Toward best practice. *American Journal of Criminal Justice*, 44, 499 – 518.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, 14(3), 399 – 441.

Rappert, B., Wilson-Kovacs, D., Wheat, H., & Leonelli, S. (2022). Evincing offence: How digital forensics turns big data into evidence for policing sexual abuse. *Engaging Science Technology and Society*, *8*(3).

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907 – 39. *Social Studies of Science*, 19(3), 387 – 420.

Straus, S. E., Tetroe, J., & Graham, I. D. (eds.) (2013). *Knowledge Translation in Health Care: Moving from Evidence to Practice.* London: John Wiley & Sons.

Telep, C. W., & Winegar, S. (2015). Police executive receptivity to research: A survey of chiefs and sheriffs in Oregon. *Policing*, 10, 241 – 249.

Wæraas, A., & Nielsen, J. A. (2016). Translation theory 'translated': Three perspectives on translation in organizational research. *International Journal of Management Reviews*, 18(3), 236 – 270.

Wilson-Kovacs, D. (2021). Digital media investigators: Challenges and opportunities in the use of digital forensics in police investigations in England and Wales. *Policing: An International Journal*, 44(4), 669 – 682.

Wilson-Kovacs, D. (2022). Deliberating forensic genetics innovations: The case of rapid DNA technologies in England and Wales. In V. Toom, M. Wienroth, & A. M'charek (eds.), *Law, Practice and Politics of Forensic DNA Profiling: Forensic Genetics and their Technolegal Worlds* (pp. 111 – 126). London: Routledge.

Wilson-Kovacs, D., & Wilcox, J. (2023). Managing policing demand for digital forensics through risk assessment and prioritization in England and Wales. *Policing: A Journal of Policy and Practice*, 17, paac106.

Woolgar, S., Coopmans, C., & Neyland, D. (2009). Does STS mean business? *Organization*, 16(1), 5 – 30.

Wyatt, D. (2014). Practising crime scene investigation: Trace and contamination in routine work. *Policing and Society*, 24(4), 443 – 458.

Sandra Walklate

56 Victimization

Abstract: This chapter, in using the two main approaches found within victimology (one focusing on interaction processes, the other on structural processes) considers the extent to which a case can be made for a digital victimology. It concludes that whilst the digital world poses many challenges in relation to victimization there is much to be gained from considering the continuities between the pre-digital and the post-digital world in understanding such victimization.

Keywords: victimology, spacelessness, continuities, algorithmic victimization, social media, the Internet of Things

Introduction

The study of victimization is largely associated with the sub-discipline of criminology labeled by Mendelsohn post the Second World War as victimology. Since that time different theoretical perspectives have emerged under this heading all making differently informed claims about suffering as a result of criminal behavior: the focal concern of victimology (Walklate, 2023). Thus, in many ways understanding the relationship between suffering—a key focus for studies of victimization—and crime is a concern which gained momentum in the 1960s as a result of the growth and development of the criminal victimization survey alongside the emergence of victim support organizations. As the nature of the crime problem has evolved along with the evolution of societies, so has the nature of criminal victimization. This is especially notable as societies have shifted from being pre-digitized to digitized. During this process of change the term victimization has been used in two main ways in trying to understand the impact that criminal victimization has on people.

The first of these considers victimization as the product of an interactional process. This focuses attention on what takes place between people resulting in one person (or persons) acquiring the status of being a victim (of crime). Rock (2002) suggests that understanding victimization in this way focuses attention on not only the process of interaction that takes place between the victimized and the victimizer but also how and why people come to see themselves as a victim, what this means to them, what it means when this status is recognized (or not) by others, and under what circumstances all of these become problematic. This understanding of victimization particularly addresses the question of how it is that some people embrace a victim identity and others do not. This process is still not fully understood by victimologists. The second understanding of victimization draws attention to, not so much to the interactional processes of becoming a victim, but more to the ways in which such experiences are not evenly distributed throughout any society. The criminal victimization survey has proved to be a powerful instrument in understanding this patterning of victimization. As a result of

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-057

criminal victimization survey data, it is widely acknowledged that being a victim of crime in a diverse range of jurisdictions is mediated by variables such as age, sex, gender, ethnicity, and so on.

Importantly these two ways of thinking about victimization do not necessarily align with dictionary definitions of this word. Dictionary definitions tend to be more generic (focusing on the harm and/or suffering experienced by people through no fault of their own and not always connected with actions defined as criminal). Nevertheless, these different understandings of victimization reflect differing ways of gathering data about such victimization. The first being more qualitative in emphasis the second being more quantitative. Moreover, the different theoretical perspectives available within victimology can also elide and conflate the harm done as a result of criminal victimization in different ways (see Walklate, 2023). However, for the purposes of this discussion the two understandings outlined above provide some valuable insight into the contemporary strengths and weaknesses of thinking about victimization in relation to the digital world: victimization as a process or victimization as a structurally informed experience.

Victimization and digital criminology

In the mid to late 1960s when victimization (from crime) first came into view, the nature and extent of a digitized society was, at most, embryonic. These early concerns with criminal victimization reflected what were then relatively conventional understandings of crime drawing attention to the harm done by street crime (in public) and the harm done by criminal behavior occurring in the home (in private). To a lesser extent some of this work also addressed the harm done by corporate crime. In all these settings the focus was on experiences occurring in offline settings. Since that time the presence of the internet and its powerful role in everyday life has grown exponentially. This change has been so all-embracing that from the viewpoint of the 21st century Harris (2018) argues that criminal victimization is characterized by spacelessness. In other words, the perpetration of such victimization is committed across time and space in such a way as it becomes inescapable for those so victimized. This is especially the case when such victimization pertains to the gendered nature of different kinds of violence(s). Whilst Harris developed this term to capture women's experiences of partner abuse in all its forms, it is especially designed to capture the essence of a wide range of harmful behaviors perpetrated on women through the medium of online environments by men known and unknown to them (see Abuse by McAlinden). For example, a woman living in Australia may remain subjected to such violence(s) by an (ex)-partner living in a different time zone and geographical space. Developing an understanding of digital victimization further, Stratton et al. (2016) suggest there are three dimensions through which this kind of victimization occurs and can be understood: through the impact of social media, the role of big data, and the Internet of Things.

Using these three dimensions as a heuristic and illustrative device informs the following discussion of digital victimization.

Digital victimization and social media

The presence and impact of social media platforms of all kinds has, for many people across the globe, become an intrinsic part of everyday life (see Social Media by Twigt). These various platforms have facilitated means of immediate communication not only between families and friends but also for a wide variety of other social groups. In so doing they afford the same and/or similar possibilities for harm (victimization) as can be experienced face-to-face: bullying, verbal abuse, fraud, stalking, harassment, and so on. In some circumstances (for example for young adults) the line drawn between their online and offline experiences of this kind of victimization is barely visible and is hardly articulable. In many ways their online and offline lives are so intertwined they are as one. Thus, the challenge for criminologists and victimologists interested in exploring the impact of this kind of digital victimization are profound. The challenges require an understanding the complex interconnectedness of these experiences and pose similar challenges for criminal justice professionals in evidencing such victimizations though mobile phone apps, text messaging, and so on. Simultaneously these same social media platforms have also become an increasingly powerful presence, used by victims themselves, to call 'offenders' to account. This has been powerfully epitomized by the #MeToo movement (though there are many others) and can be thought of as empowering (though see Cossins, 2020). Such movements can have ambivalent consequences. They might afford moments in which unheard voices are heard but equally the capacity for such movements to criminalize those not established as criminal are in law are profound (Walklate, 2020). Moreover, such victims can themselves lose further control of their experiences as their stories are taken up by others which can result in further victimization (Wood et al., 2019) and the creation of new forms of victimhood. This brief overview implies that the victimization experienced through the medium of social media parallels the two understandings of victimization outlined at the beginning of this discussion. The role of big data and its capacity for victimization offers a different lens on such experiences.

Digital victimization and big data

Big data impacts upon crime and justice processes primarily through the medium of risk assessment practices (see Big Data by Završnik and Sentencing and Risk Assessments by Ugwudike). Such practices are not new since decision-making about repeat and/or habitual offenders has always been informed by the extent to which past behavior might predict future behavior. Whilst such tools are many, varied, and carry different labels (clinical, actuarial, and algorithmic), big data has been particularly powerful

in informing the last of these approaches: the algorithmic (see Algorithm by Leese). Algorithmic risk assessment tools, deploying machine learning techniques, use a wide range of variables in complex models to inform decision-making. When used as predictive tools in criminal justice practice, they focus attention on 'risky' offenders or 'at risk' victims and presume that people do not change (see also Mythen, 2014). Importantly the development of such tools has been intrinsic to the growth of the digital society above and beyond criminal justice. However, in assuming that people do not have the capacity to change, they presume at least two things: that the relationship between the victim and offender is static and dichotomous (failing to recognize the blurred boundaries between these two terms and people's real lives) and that such tools are objective. These assumptions bring into view the problematic questions posed when *offenders are also victims and vice versa.* Thus, whilst on the one hand the big data driven capacities of such digitally informed tools suggest the capability of objectively informed decision-making, such a presumption can open the door to further victimization when a closer look is taken on how they are built and used (see especially, D'Ignazio and Klein, 2023).

For example, Werth (2017: 809) suggests (in discussing parole decisions) "… field personnel devalue actuarial tools and instead privilege experiential expertise and moral judgement of personhood" with Walklate and Mythen (2011: 109) arguing that it is possible "… risk, and those deemed at risk, are not forensically measured at all: they are constructed within a logic of norms and values *that are felt*" (my emphasis). In other words, professionals make judgments based on their feelings and intuition, in the context of the case with which they are dealing. Berk (2017) goes on to state that from bail decisions to parole decisions algorithmic risk assessments must be a demonstrable improvement of human decision-making but also adds that the bar for such improvement is not remarkably high. Importantly then algorithms are not necessarily devoid of bias, and such biases can be compounded by the statistical techniques which underpin these tools (see Sentencing and Risk Assessments by Ugwudike). Thus, as Hannah-Moffat (2019: 460) points out, "… big data technologies, like risk instruments, simultaneously appear neutral and authoritative, which can make them powerful tools of governance." Yet increasingly much decision-making within criminal justice is being made informed by such invisible tools and in doing so, alongside the implied lack of accountability associated with such invisibility, can add to the harms they do especially in relation to offenders. Such developments bring into view offenders as victims and thereby offers a challenge to the unidimensional understandings of victimization with which this discussion began. The final strand of digital criminology of interest to the discussion of victimization is the increasing presence of the 'Internet of Things' frequently referred to in shorthand as IoT (see Internet of Things by Milivojevic).

Digital victimization and the Internet of Things

IoT denotes objects with the capacity for processing data and interacting with other similarly equipped objects over the internet. Such objects, often referred to as 'smart,' include mobile phones, fire alarms, door locks, fitness trackers, heating systems, refrigerators, personal assistants (like Alexa and Cortana), and so on. Put simply such devices can perform a range of tasks, can be operated remotely, and can interact with one another. They also afford further opportunities for criminal activity (through hacking for example) and whilst offering significant opportunities for increased surveillance (in being able to observe home deliveries and other domestic activities remotely) such increased surveillance is also an opportunity for abuse.

Perpetrators of domestic abuse have not been slow to recognize the opportunities provided by the increasing presence of smart technologies to engage in the further abuse and control of their victims: bullying, stalking, tracking, sharing intimate photographs without permission with others, controlling the heating system. When these behaviors are taken together they amount to what Harris and Woodlock (2019) and others have termed digital coercive control. Importantly this kind of abuse and control transcends space as Harris (2018) has pointedly argued both its perpetration and its impact on those so victimized. This adds a further dimension to understanding how those experiencing such victimization can or cannot escape from them. It is a particularly pertinent observation for those living with the challenges of mental and/or physical disabilities who may already be totally dependent on their carer and whose experience of abuse can be exacerbated in the abusive use of such technology. There are of course many other structural features of victimization (like age, ethnicity, and so on) which also inform how victimization through the IoT might be experienced. So whilst IoT may have been embraced as a positive development for everyday life its conveniences notwithstanding, the opportunities for the further perpetration of victimization by these means have added to those same experiences of victimization pre-existing IoT and arguably worsened by it. In other words, there is much continuity here between the pre- and post-digital setting significantly illustrating the ongoing importance of feminist informed continuum thinking (Kelly, 1988).

Overview

The advancements made with increasing digitization in society present both opportunities and risks. In terms of experiences of victimization there is arguably more continuity with the past than perhaps the proponents of digital developments have as yet recognized. This is certainly evident in the discussion of the three strands of digitization discussed above which have made their presence felt in criminology. Hence, from a point of view, the concept of victimization itself understood as a product of an interactional process and/or as a structural feature of any social system, remain relatively untarnished. At the same time, it is also possible discern that in some respects

the presence of digital victimization has also blurred the boundaries between structural and interpersonal victimization in significantly nuanced ways for many people but especially those living with violence(s). Thus a woman living with abuse both digitally and face-to-face, finds herself in a very challenging set of circumstances indeed. There is more scope for a digitally informed criminology-victimology to explore these interconnections.

It is also possible to discern the risks associated with calling perpetrators to account through social media campaigns for both victims and perpetrators. This is especially the case when those named as perpetrators are subsequently found not guilty. Moreover, the dangers inherent in the bias's evident in algorithmic informed judgments have also been well-documented (see inter alia Berk, 2017). At the same time, of course, the presence of 'strategic witnessing'—the use of mobile phone footage as evidence especially in relation to mass events of violence (Ristovska, 2016), alongside the wider availability of digitally based technology, has improved some aspects of criminal justice delivery. For example, the use of virtual platforms for multi-agency risk assessment meetings for cases of domestic abuse became an important and ongoing development for police forces in England and Wales during the COVID-19 pandemic (Walklate et al., 2021). However, the jury remains out on the value of online spaces for court processes as a whole (see for example Harker and Ryan, 2022 on the family court and Rossner and Tait, 2023 on the criminal court.). Again, there is considerable space here for criminologists and victimologists to explore more deeply what kind of impact these developments may have on people's experiences of victimization.

Conclusion

From this brief overview of some of the newer and continuing features of victimization in the 21st century it is possible to discern the considerable contribution that has yet to be made in understanding the nature and impact of digitization on people's experiences of victimization on a routine daily basis.

- Whilst considerable in-roads have been made in understanding the impact of the digital on a wide range of abusive behavior on women and other groups, the nature and extent of that impact has yet to be fully realized for those whose experiences of such victimization are a result of interaction (that is generated by people who are known to them) and yet compounded by the structural (that is generated systemically through processes of marginalization). This observation is not a new one.
- There is much here that reveals continuities between the past and the present. For example, whilst social media and IoT may be a key means by which stalking, harassment, bullying, and so on are now perpetuated, these behaviors are not new. They have always been perpetuated if in the past by different means (landline telephones, following someone in the street, verbal abuse, and so on).

– The same might be said for the contemporary use of big data and the ways in which traditional biases built into clinical and actuarial risk assessment practices are similarly perpetuated in the use and deployment of algorithmic based practices.

Final comment and as a word of caution, the rise in digital criminology-victimology may simply reinvent victimization as though the digital world and its potential equates with new experiences of harm. Understanding people's lives as they are lived is one way to avoid such reinvention. In appreciating the continuities of the present with the past is one place to start.

Suggested reading

For those interested in the extensive nature and impact of technology facilitated victimization then the edited collection by B. Harris & D. Woodlock, *Technology and Domestic and Family Violence* (London: Routledge, 2023) is a good place to start.

On the nature, extent, and potential of digital justice the work by E. Katsh & O. Rabinovich-Einy, *Digital Justice: Technology and the Internet of Disputes* (Oxford: Oxford University Press, 2023). might be one place to start.

Introductory work specifically on cyber victimology can be found in D. Halder, *Cyber Victimology: Decoding Cyber-Crime Victimisation* (London: Routledge, 2022). This book digs a little deeper into the different facets of such victimization in an easy and accessible style.

For a more thorough introduction and discussion of victimology and victimization see S. Walklate, *Advanced Introduction to Victimology* (Cheltenham, UK : Elgar Publishing, 2023).

References

Berk, R. (2017). Using ML in criminal justice risk assessments: The frontiers of machine learning. https://www.youtube.com/watch?v=gdEPPRhNu34&app=desktop

Cossins, A. (2020). Feminist criminology in a time of 'digital feminism': Can the #MeToo movement create fundamental cultural change? In S. Walklate, K. Fitz-Gibbon, J. M. Maher, & J. McCulloch (eds.), *The Emerald Handbook of Feminism, Criminology and Social Change* (pp. 69 – 96). Bingley: Emerald Publishing.

D'Ignazio, C., & Klein, L. F. (2023). *Data Feminism.* Cambridge, MA; MIT Press.

Hannah-Moffat, K. (2019). Algorithmic risk governance: Big data analytics, race, and information activism in criminal justice debates. *Theoretical Criminology*, 23(4), 453 – 470. DOI:10.1177/1362480618763582

Harker, L., & Ryan, M. (2022). Remote hearings in family courts in England and Wales during Covid-19: Insights and lessons. *Family Court Review*, 60(2), 207 – 219. https://doi.org/10.1111/fcre.12638

Harris, B. (2018). Spacelessness, spatiality and intimate partner violence: Technology-facilitated abuse, stalking and justice. In J. Maher, S. Walklate, J. McCulloch, & K. Fitz-Gibbon (eds.), *Intimate Partner Violence, Risk, and Security: Securing Women's Lives in a Global World* (pp. 52 – 70). London: Routledge.

Harris, B., & Woodlock, D. (2019). Digital coercive control: Insights from two landmark domestic violence studies. *The British Journal of Criminology*, 59(3), 530 – 550. https://doi.org/10.1093/bjc/azy052

Kelly, L. (1988) *Surviving Sexual Violence.* Cambridge: Polity Press.

Mythen, G. (2014). *Understanding the Risk Society.* London: Palgrave.

Ristovska, S. (2016). Strategic witnessing in an age of video activism. *Media, Culture & Society*, 38(7), 1034 – 1047. https://doi.org/10.1177/0163443716635866

Rock, P. (2002). *On becoming a victim.* In C. Hoyle & R. Wilson (eds.), *New Visions of Crime Victims* (pp. 1 – 22). Oxford: Hart Publishing.

Rossner, M., & Tait, D. (2023). Presence and participation in a virtual court. *Criminology & Criminal Justice*, 23(1), 135 – 157. https://doi.org/10.1177/17488958211017372

Stratton, G., Powell, A., & Cameron, R. (2017). Crime and justice in digital society: Towards a 'digital criminology'? *International Journal for Crime, Justice, and Social Democracy*, 6(2), 17 – 33. DOI:10.5204/ijcjsd.v6i2.355

Walklate, S. (2020). Living in La La Land: 'Snowflakes', social change and alternative responses to sexual assault. In P. Carlen & A. L. Franca (eds.), *Justice Alternatives* (pp. 175 – 184). London: Routledge.

Walklate, S. (2023). *Advanced Introduction to Victimology.* Cheltenham, UK: Elgar Publishing.

Walklate, S., & Mythen, G. (2011). Beyond risk theory: Experiential knowledge and 'knowing otherwise'. *Criminology and Criminal Justice*, 11(2), 99 – 113.

Walklate, S., Godfrey, B., & Richardson, J. (2021). Innovating during the pandemic? Policing, domestic abuse and multi-agency risk assessment conferencing (MARACs). *Journal of Adult Protection*, 23(3), 181 – 190. https://doi. org/10.1108/jap-11 – 2020 – 0047

Werth, R. (2017). Individualizing risk: Moral judgement, professional knowledge and affect in parole evaluations. *British Journal of Criminology*, 57(4), 808 – 827.

Wood, M., Rose, E., & Thompson, C. (2019). Viral justice? Online justice-seeking, intimate partner violence and affective contagion. *Theoretical Criminology*, 23(3), 375 – 393.

Sofia Ranchordas and Malou Beck

57 Vulnerability

Abstract: This chapter explores definitions and theories of vulnerability, emphasizing its complexity, elusive nature, and the impact of digitalization on exacerbating vulnerabilities in the public sector and the criminal justice system. It examines a conceptual understanding of administrative vulnerability and draws on existing scholarship theorizing vulnerability as exposure to harm, individual and particular, universal and inherent, and multi-layered.

Keywords: vulnerability, digitalization, criminal justice, public sector governance, inequality

Introduction

The academic conceptualization of vulnerability first emerged in the early 1980s in the study of natural disasters (Delor and Hubert, 2000: 1561). Today, vulnerability is explored in multiple sciences and is used in both the analog and digital contexts. In cybersecurity, any flaw in an organization's internal controls, system procedures, or information systems can be regarded as a 'vulnerability.' Cybersecurity vulnerabilities affect individuals, businesses, and nation states, and research has shown that some countries (e. g., India and South Korea) are particularly vulnerable to cyber threats (Subrahmanian et al., 2015). Similarly, in organizational studies, vulnerability has also been connected to the identification of threats and risks within corporations. However, in this field of study, vulnerability has also been approached from a management perspective: originally regarded as a weakness, vulnerability in management equates nowadays to empathic and humble leadership (Brown, 2018).

In law and criminology, vulnerability is primarily associated with features of specific groups. For instance, persons with disabilities, migrants, or children are commonly considered vulnerable. The subfield of zemiology draws attention to the meaning of vulnerability in relation to social harms, identifying structural disadvantages, systemic injustices, and unequal power structures (Canning and Tombs, 2021). In law, vulnerability is traditionally connected to belonging to a certain group (e. g., children) and as such, vulnerable citizens are awarded additional legal protection when they cannot exercise their rights. Vulnerability theories developed in the context of legal feminist studies depart from this position. They challenge the existence of 'an average citizen' free from vulnerability. Instead, Fineman (2008), for example, suggests that vulnerabil-

Disclaimers and acknowledgments: This research was funded by the Netherlands Organization for Scientific Research (NWO) in the context of the NWO-Vidi Project 'Vulnerability in the Digital Administrative State' led by Sofia Ranchordas.

Open Access. © 2025 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. https://doi.org/10.1515/9783111062037-058

ity is universal and inherently human. Thus, all individuals can be vulnerable due to an ever-present risk of being harmed: what varies is a person's degree of vulnerability. Interestingly, this position is mirrored in the etymology of the term 'vulnerable.' Originating from the Latin term *vulnus* or wound (Malgieri and Niklas, 2020: 3), vulnerability depicts "the fragile and contingent nature of personhood" (Beckett, 2006: 3). To illustrate, senior citizens with limited digital literacy may be vulnerable online due to their higher risk of being targeted by cybercriminals (see Cybercrime by Holt and Holt). However, young people who have profound digital skills may also encounter moments of digital vulnerability. This can occur when they face personal challenges, such as a traumatic life event that temporarily reduces their ability to navigate complex digital tools independently (Calderón Gómez, 2019). While consensus exists on vulnerability as the susceptibility to be harmed, the broader aspects of the concept remain elusive.

This contribution explores key definitions and theories of vulnerability and their significance in the digital context. In criminology, vulnerability provides a valuable tool to reveal and attend to social inequalities and unjust societal structures. In addition, the concept sheds light on the human dimension in today's digital society.

Theories of vulnerability: the state-of-the-art

While progress has been made in protecting vulnerable groups, the definition of vulnerability remains a fragmented exercise (Schroeder and Gefenas, 2009). Definitions of vulnerability can be vague and biased, especially when tied to particular identity groups. Moreover, the term 'vulnerability' is often misused in academia, popular science, and media, jeopardizing its intended meaning and protective value. Four conceptualizations of vulnerability stand out, namely (i) exposure to harm; (ii) individual particularities; (iii) the universal character of vulnerability; and (iv) the existence of multiple layers of vulnerability.

First, vulnerability can be defined by reference to being "exposed to the possibility of being attacked or harmed, either physically or emotionally" (Schroeder and Gefenas, 2009: 114). This implies that once an individual has been harmed, they are no longer vulnerable. While exposure to harm is the external dimension of vulnerability, the inability to defend oneself against harm its the internal element (Delor and Hubert, 2009). For instance, citizens with limited digital literacy may be more exposed to digital harms such as identity theft. As such their internal vulnerability manifests itself in the reduced capacity to protect their personal data against such this type of cybercrime.

A second strand of scholarship associates vulnerability with *particularity*, qualifying individuals as vulnerable based on specific situations and socio-economic contexts. Despite being widespread in legal and criminology scholarship, the *particular* approach to vulnerability has been critiqued due to the risk of profiling, stigmatization, and grouping people as vulnerable (Heri, 2021). This has been common in the context of hate crimes which focuses on the collective experiences of minority groups and their

victimization (Chakraborti and Garland, 2012). While this particularity-qualification is practical from a policy perspective, this label does not account for the complexity and fluidity of vulnerability, individual needs, and may thus "not reliably protect many individuals from harm" (Levine et al., 2004: 44). This is particularly true in the digital realm, where data pertaining to individuals within the same population groups may be categorized differently, despite their shared vulnerabilities (Malgieri and Niklas, 2020).

The third set adopts a ***universal*** approach to vulnerability, embracing Martha Fineman's theory of the ubiquitous character of vulnerability (Fineman, 2008, 2010, 2019a). Fineman argues that human vulnerability is defined by universality and is "inherent to the human condition" (Fineman, 2008: 1). Fineman's feminist legal notion of a *universal* vulnerable legal subject assumes that vulnerability is not centered around specific groups, rejecting the understanding that vulnerability can be classified based on individuals' particularities such as race, gender, or ethnicity (Fineman, 2010: 253). Scholarship has interpreted the departure from vulnerability as a categorization of certain 'weak' groups or 'victims' as a positive development in the combat against the stigma of the term (Mackenzie, 2014; Cole, 2016; Gilson, 2011; Malgieri and Niklas, 2020; Kohn, 2014). In Fineman's work, being human means being vulnerable. Vulnerability is converted into a feature that cannot be eliminated and is often beyond human control (Fineman 2019b: 53; Fineman, 2008: 9). Thus, Fineman defines vulnerability as "the unavoidable susceptibility to change, both positive and negative, in our physical wellbeing that we experience over the life course"(Fineman 2019b: 57). Moreover, Fineman advocates for a responsive state, criticizing today's neoliberal state which places excessive focus on the liberty and autonomy of the individual forgetting human dependency and vulnerability (Fineman, 2010). In the digital context, such universal approach to vulnerability entails an acknowledgement that all citizens can be harmed by the digitalization of law enforcement or the criminal justice system. Thus, vulnerability can also be inflicted on individuals that fall out of common vulnerability categories. For instance, vulnerability arising from cyber threats, such as online scams or e-mail phishing, is not limited to specific population groups.

The fourth strand of scholarship regards vulnerability as a complex, fluid, and multi-layered concept, thus building upon universal perceptions but refining it by explaining how some individuals may be confronted with more layers of vulnerability than others (Luna, 2009: 128). Luna's notion of vulnerability rejects the stigmatization and simplistic categorization of individuals in vulnerable groups (Luna, 2009: 124). To illustrate, being a single mother of four children can constitute a layer of vulnerability. Low income and health problems, when present, will add layers to the vulnerability experienced by this mother, leading to 'cascade effects' (Luna, 2019: 91). The existence of multiple cumulative layers of vulnerability is often considered in sentencing in criminal law. The assessment of the presence of these layers is thus relevant to evaluate the vulnerability of individuals in criminal investigations.

Sources of vulnerability

Scholarship on vulnerability has sought to understand vulnerability through its many *sources* and *states.* Vulnerability can be permanent (e.g., physical disabilities) or temporary (e.g., a life event such as the death of a loved one). Vulnerability sources can also be inherent or situational: *inherent sources* of vulnerability pertain to the human condition (age, gender, health), while *situational sources* are more context-specific and often "caused or exacerbated by social, political, economic or environmental factors" (Mackenzie, 2014: 39). These sources of vulnerability do not exist in isolation, as they can be closely intertwined and can reinforce or trigger one another (Mackenzie, 2014). For example, individuals experiencing poor mental health may be more vulnerable to emotional distress from online harassment. Moreover, inherent and situational sources of vulnerability can be temporary or long-lasting and can take dispositional or occurrent forms (Mackenzie, 2014: 39). In addition, part of the situational vulnerability sources are pathogenic vulnerabilities, which are inflicted through structural disadvantage and thus result from "and from social domination, oppression, or political violence" (Mackenzie, 2014: 39). Such pathogenic vulnerabilities can, for instance, occur in discriminatory and racially biased algorithms in criminal sentencing (see Sentencing and Risk Assessments by Ugwudike). When such algorithms mirror long standing structural disadvantage and unequal power structures, they give rise to pathogenic vulnerabilities for such oppressed population groups.

Our position and administrative vulnerability

While there are various perspectives on vulnerability, we adopt a socio-legal approach and understand vulnerability as a universal, individual, and multi-layered concept. This approach matches partially with vulnerability concepts in the field of digital criminology. Vulnerability is the susceptibility of being placed in a position of economic, social, ecological, or legal disadvantage with potential harm as the result. In our administrative law and sociolegal scholarship, we develop the term *'administrative vulnerability'* to convey the inability to critically engage with digital government, bureaucracy, and exercise rights before government on equal terms. Administrative vulnerability is a fluid concept that encompasses intrinsic and situational vulnerability. It conveys a state of inequality, i.e., citizens experiencing situational vulnerability will not be able to exercise their rights on equal terms. Administrative vulnerability is problematic for four reasons: first, citizens have rights before government and are entitled to receiving different types of assistance and public services; second, citizens typically do not have choice when it comes to the provision of public services as in many countries, the state is their sole provider; third, the relationship between state and citizens is inherently defined by power asymmetries; fourth, in a democratic society governed by the rule of law, citizens experiencing vulnerability can rightfully expect to be protected

rather than be harmed by government. Nowadays, public services and public assistance such as social welfare benefits are provided mainly through digital government in numerous high-income countries. For instance, citizen will be regarded as administratively vulnerable if digital government platforms do not provide easy access to such services.

Significant power asymmetries between citizens and governments typically define government transactions, particularly criminal investigations and the gathering of digital evidence. However, some situations will place individuals in more vulnerable conditions than others, depending on the existing layers of vulnerability, structural disadvantage, and the specific relationship between citizens and the government. Furthermore, administrative vulnerability can occur in many different contexts of digital government, ranging from accessing and engaging with digital services to becoming victims to algorithms or automated decision-making systems. For instance, if individuals are unable to engage with digital tools, they can become vulnerable when unable to challenging decisions taken against them and, thus, accessing justice. Moreover, sentencing algorithms, surveillance systems and policing softwares can give rise to vulnerabilities if, for instance, racially biased learning data is used, ultimately resulting in discriminatory outcomes and fundamental rights violations (see Bias by Oswald and Paul).

While it is true that everyone can be vulnerable, not all citizens are susceptible to experience vulnerability at the same level. Our position acknowledges that some individuals are predisposed to particular layers of vulnerability because of endogenous or exogenous sources of vulnerability or risk factors. For example, individuals with limited digital literacy may be more susceptible to different forms of cybercrime such as phishing scams, identity theft, or online fraud. The cumulation of layers of vulnerability should also be considered as it can have de facto a punitive effect on citizens, underlining the urge to draw attention to it in critical criminology scholarship. For example, welfare recipients who experience multiple layers of vulnerability (low income, criminal record, low literacy, and limited digital skills) may be at higher risk of being targeted by welfare investigations due to algorithmic biases. At the same time, they may inadvertently confirm this risk factor by making unintentional mistakes when filling in forms or providing inaccurate information to the government due to their limited knowledge and skills. As the state unintentionally but directly inflicts administrative vulnerability, this concept also acknowledges Fineman's idea that the state should be responsible for combatting and accounting for it.

Critical criminology scholarship would benefit from a layered approach and the departure from negative and stigmatizing perceptions of vulnerability that label entire population groups as vulnerable. These perceptions do not always entail a critical inquiry into the experience of these populations and how they may be stigmatized because of the 'vulnerability-label.' Instead, all individuals are prone to administrative vulnerability and carry layers that can make them susceptible to it. This underlines the universal and complex nature of the concept as defined in Fineman's vulnerability theory.

Vulnerability, criminology, and the digital

With the growing use of technology (e. g., biometrics, big data, AI) in predictive policing, welfare fraud investigations, and criminal investigations, some groups have become more vulnerable to biases and the risk of wrong accusations (Susser, 2021). The use of biometrics to govern, regulate, and police by identity has, for instance, proven problematic to ethnic minorities (Skinner, 2020). Ethnic minorities are often discriminated against by predictive systems that score them negatively, suggest they will commit crimes and welfare fraud more often, and surveil them more closely than other segments of the population (Ferguson, 2017: 1109). Biometric identification typically affects these groups negatively as these groups are more regularly selected for random biometric identification at airports, train stations, and other public spaces (Magnet, 2011). The compromise text of the AI Act which will regulate a wide range of AI systems in the European Union, aims to impose future limits on real-time biometric identification in public spaces. This is crucial to avoid that ethnic minorities experience additional vulnerability in criminal investigations, particularly when there have been several documented cases of wrongful incrimination. However, a more profound solution here would entail the elimination of unjust power and opportunity structures and structural disadvantage to ethnic minorities as the root cause of biased datasets and vulnerability.

Furthermore, debates surrounding the emergence of e-courts and a digitalized criminal justice system have created new digital vulnerabilities for various groups (see Online Courts by Mentovich and Rabinovich Einy). While traditionally, there has been a dichotomy between the victim as the vulnerable and the offenders as the party inflicting vulnerabilities on individuals (Green, 2012), recent digital advancements have urged us to rethink such narrow understandings of vulnerability. For instance, the use of audio-visual communication tools in the context of courts has given rise not only to new vulnerabilities for victims, their networks, and crime witnesses but also to offenders and incarcerated individuals. For instance, Parsons and Sherwood (2016) claim that communication and information provision can be challenging for detained individuals with learning disabilities. This can have far-reaching consequences and even lead to detrimental miscarriages of justice. While such issues are already evident in paper-based communication channels (Parsons and Sherwood, 2016: 569), digitalizing such processes risks exacerbating such vulnerabilities.

In addition, courts have not only given rise to vulnerability related to access to justice, but the use of remote access technologies in courts also raises questions regarding power relations, justice, and fairness (McKay, 2022: 74). Vargas et al. (2019) argue that citizens' e-contact with the criminal justice system has given rise to new digital vulnerabilities through the risk of data breaches in the context of police-dispatcher radio communication, primarily affecting racial minorities in disadvantaged neighborhoods. Thus, "citizen vulnerability occurs not only physically during police stops on the streets or country courts, but also digitally across cyberspace and radio waves" (Vargas et al.,

2019: 74). Notwithstanding, in the criminal justice system, individuals generally find themselves in vulnerable positions, which can be exacerbated through digital tools. Prevailing attributes that render individuals vulnerable in bail decisions are race, age, sex, and socioeconomic status (Hughes et al., 2022). In conclusion, the digital dimension of law enforcement and judicial procedures exposes individuals to additional vulnerabilities and thus potential harm.

Conclusion

Vulnerability is a complex, universal but also individual, multi-layered concept. To be human is to be vulnerable and each and every one of us can be vulnerable in our own way. While there has been progress in the protection of vulnerable groups, commonly used definitions of vulnerability remain restricted by their connection to closed groups and categories. This contribution highlighted four central conceptualizations of vulnerability based on the notions of (i) exposure to harm; (ii) individual particularities; (iii) the universal character of vulnerability; and (iv) the existence of multiple layers of vulnerability. Furthermore, we explained that vulnerability stems from different sources, such as inherent, situational, or pathogenic sources, and can manifest itself in occurrent, dispositional, temporary, or permanent form (Mackenzie, 2014). While the digitalization and automation of the public sector and criminal justice system has given rise to new types of vulnerabilities, such as identity fraud, problematic digital evidence practices, cybercrimes, predictive fraud risk algorithms, or access to e-courts, these conceptualizations of vulnerability sources remain relevant and applicable. We argued that vulnerability is a *state* that every individual can experience, rather than a *status* or label that is given to underprivileged groups. While some individuals are more exposed to different layers of vulnerability and may thus require additional protection against for example cybercrime or online harassment, the state should be mindful that vulnerability is an inherent feature of human identity.

Moving forward, scholars, policymakers, and governments should be attentive to the impact of the digital transformation on individuals. The digital creates a context in which existing layers of vulnerability become more salient and new layers such as digital and algorithmic illiteracy emerge.

Suggested reading

Dehaghani, R., Fairclough, S., & Mergaerts, L. (eds.) (2023). *Vulnerability, the Accused, and the Criminal Justice System: Multi-jurisdictional Perspectives.* London: Routledge.

Fineman, M., & Spitz, L. (eds.) (2024). *Law, Vulnerability, and the Responsive State Beyond Equality and Liberty.* London: Routledge (forthcoming)

Green, S. (2012). Crime, victimisation and vulnerability. In S. Walklate (ed.), *Handbook of Victims and Victimology* (pp. 107–134). Cullompton: Willan

Hickey, J. (ed.) (2023). *The Foundations of Vulnerability Theory: Feminism, Family, and Fineman* (1st ed.). London: Routledge.
Malgieri, G. (2023). *Vulnerability and Data Protection Law.* Oxford: Oxford University Press.
Taylor, S. (2022). 7: The problem with the current agenda: Focus on vulnerability. In *Hate Crime Policy and Disability* (pp. 144 – 167). Bristol: Bristol University Press. Available from: < https://doi.org/10.51952/9781529217902.ch007> [Accessed 03 August 2023]

References

Beckett, A. (2006). *Citizenship and Vulnerability: Disability and Issues of Social and Political Engagement.* Basingstoke: Palgrave MacMillan.
Brown, B. (2018). *Dare to Lead: Brave Work. Tough Conversations. Whole Hearts.* London: Random House.
Calderón Gómez, D. (2019). Technological capital and digital divide among young people: An intersectional approach. *Journal of Youth Studies*, 22(7), 941 – 958.
Canning, V., & Tombs, S. (2021). *From Social Harm to Zemiology: A Critical Introduction.* London: Routledge.
Chakraborti, N., & Garland, J. (2012). Reconceptualizing hate crime victimization through the lens of vulnerability and 'difference'. *Theoretical Criminology*, 16(4), 499 – 514.
Cole, A. (2016). All of us are vulnerable, but some are more vulnerable than others: The political ambiguity of vulnerability studies, an ambivalent critique. *Critical Horizons*, 17(2), 260 – 277
Delor, F., & Hubert, M. (2000). Revisiting the concept of 'vulnerability'. *Social Science & Medicine*, 50(11), 1557 – 1570.
Ferguson, A. G. (2017). Policing predictive policing. *Washington University Law Review*, 94(5), 1109 – 1189.
Fineman, M. A. (2008). The vulnerable subject: Anchoring equality in the human condition. *Yale Journal of Law & Feminism*, 20, 1 – 24.
Fineman, M. (2010). The vulnerable subject and the responsive state. *Emory Law Journal*, 60, 251 – 267.
Fineman, M. A. (2019a). Vulnerability and social justice. *Valparaiso University Law Review*, 53, 341 – 369
Fineman, M. A. (2019b). Vulnerability in law and bioethics. *Journal of Health Care for the Poor and Underserved*, 30(5), 52 – 61.
Gilson, E. (2011). Vulnerability, ignorance, and oppression. *Hypatia*, 26(2), 308 – 332.
Green, S. (2012). Crime, victimisation and vulnerability. In S. Walklate (ed.), *Handbook of Victims and Victimology* (pp. 107 – 134). London: Routledge.
Heri, C. (2021). *Responsive Human Rights: Vulnerability, Ill-treatment and the ECtHR.* Oxford: Hart Publishing.
Hughes, D., Colvin, E., & Bartkowiak-Théron, I. (2022). Police and vulnerability in bail decisions' *International Journal for Crime, Justice and Social Democracy*, 11(3), 122 – 138.
Kohn, N. A. (2014). Vulnerability theory and the role of government. *Yale Journal Law & Feminism*, 26, 1 – 28.
Levine, C., Faden, R., Grady, C., Hammerschmidt, D., Eckenwiler, L., & Sugarman, J. (2004). The limitations of "vulnerability" as a protection for human research participants. *The American Journal of Bioethics*, 4(3), 44 – 49.
Luna, F. (2009). Elucidating the concept of vulnerability: Layers not labels. *IJFAB: International Journal of Feminist Approaches to Bioethics*, 2(1), 121 – 139.
Luna, F. (2019). Identifying and evaluating layers of vulnerability – a way forward. *Developing World Bioethics*, 19(2), 86 – 95.
Mackenzie, C. (2014). The importance of relational autonomy and capabilities for an ethics of vulnerability. In C. Mackenzie, W. Rogers, & S. Dodds (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy* (pp. 33 – 59). New York: Oxford University Press.
Magnet, S. (2011). *When Biometrics Fail: Gender, Race, and the Technology of Identity.* Durham, NC: Duke University Press.

Malgieri, G., & Niklas, J. (2020). Vulnerable data subjects. *Computer Law & Security Review*, 37, 105415.

McKay, C. (2022). Remote access technologies, clinical evaluations of people-in-prison and digital vulnerability. *Recht der Werkelijkheid /Journal of Empirical Research on Law in Action*, 43(2), 68 – 88.

Parsons, S., & Sherwood, G. (2016). Vulnerability in custody: Perceptions and practices of police officers and criminal justice professionals in meeting the communication needs of offenders with learning disabilities and learning difficulties. *Disability & Society*, 31(4), 553 – 572

Schroeder, D., & Gefenas, E. (2009). Vulnerability: Too vague and too broad? *Cambridge Quarterly of Healthcare Ethics*, 18(2), 113 – 121.

Skinner, D. (2020). Race, racism and identification in the era of technosecurity. *Science as Culture*, 29(1), 77 – 99.

Subrahmanian, V. S., Ovelgonne, M., Dumitras, T., & Prakash, B. A. (2015). *The Global Cyber-Vulnerability Report.* Cham: Springer.

Susser, D. (2021). Predictive policing and the ethics of preemption. In B. Jones & E. Mendieta (eds.), *The Ethics of Policing: New Perspectives on Law Enforcement* (pp. 268 – 292). New York: NYU Press.

Vargas, R., Preto-Hodge, K., & Christofferson, J. (2019). Digital vulnerability: The unequal risk of e-contact with the criminal justice system. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(1), 71 – 88.

List of contributors

Catherine Adams is a Professor of educational computing at the University of Alberta, Faculty of Education. Her research investigates digital technology integration in educational environments, focusing on its ethical and pedagogical implications.

Anthony Amicelle is Associate Professor at Sciences Po Bordeaux, Department of Political Science, and Affiliate Professor at the University of Montreal, Department of Criminology. He works on global policymaking in relation to corruption and financial crime, as well as on the role of surveillance in the policing of transnational flows.

Jan Christoffer Andersen is a Doctoral research fellow at the University of Oslo, Department of Criminology and Sociology of Law. His research explores narratives, masculinity, stigma, and the subcultural characteristics of the incel community online.

Claudia Aradau is a Professor in the Department of War Studies at King's College London. She works on the political contestation of security practices and technologies deployed to govern individuals and populations globally.

Will Arpke-Wales is a PhD candidate at the University of Melbourne, School of Social and Political Sciences. His work focuses on the intersection between technology, harm and misogyny.

Silje Anderdal Bakken is a Postdoc at the University of Oslo, Department of Criminology and Sociology of Law. She works on the role of digitalization on crime and crime control, as well as further developing online qualitative methods.

Malou Beck is a PhD candidate at Tilburg Law School, conducting socio-legal research on Digital Government and Vulnerability.

Rocco Bellanova is a Research Professor at the Vrije Universiteit Brussel (VUB), Faculty of Law and Criminology, interdisciplinary research group Law, Science, Technology and Society (LSTS). His work studies how digital data become pivotal elements in the governing of societies.

Lee A. Bygrave is a Professor at the Department of Private Law, University of Oslo, where he is Director of the Norwegian Research Center for Computers and Law. He works primarily on researching and developing regulatory policy concerning data privacy, cybersecurity, robotics, artificial intelligence, and governance of critical internet infrastructure.

Robin Cameron is a Senior Lecturer in Criminology & Justice Studies at RMIT University. His research focuses on security through an examination of gender, race, and violence in urban and online spaces.

Janet Chan is Emeritus Professor at the UNSW Faculty of Law and Justice. Her work had been focused on technological change in policing, policing training and cultures, and aspects of creativity.

Stefano Mazzilli Daechsel is a Postdoctoral fellow at the University of Oslo, Department of Criminology and Sociology of Law. His work explores the social, criminological, and philosophical implications of emerging technologies.

Katja de Vries is Senior Lecturer in public law at the University of Uppsala, Department of Law. Her field of expertise is Law and AI, where she currently specializes on the regulation of generative AI, data sharing, and the right to repair.

Heather Dewey-Hagborg is an artist and biohacker whose work engages genetics, privacy, and the social implications of technology.

Ravida Din is a PhD candidate at the University of Ottawa, Institute of Feminist & Gender Studies. She works on the role of digital technologies in gender-based violence.

Simon Egbert, Senior Researcher, Bielefeld University, Faculty of Sociology. He is interested in algorithms, platforms, and predictive analytics, particularly in the context of policing.

Katja Franko is Professor of Criminology at the University of Oslo, Norway. She has published extensively on globalization, migration and border control, international police co-operation, and the use of advanced information and communication technologies in crime control strategies.

Pete Fussey is Professor of Criminology at the University of Southampton. He focuses on human rights and emerging technology, having published widely, authored UN standards, and led national strategy on these issues.

Vasilis Galis is Professor in Science and Technology Studies at the IT University of Copenhagen. Galis was the PI for the Nordforsk funded project Critical Understanding of Predictive Policing. Galis' research focuses on the digitalization of the welfare state and law enforcement, participating in academic and public debates regarding the datafication of security, the border, and the judicial system.

Nicholas Gibbs is Assistant Professor in Criminology at Northumbria University. His work focuses on critical sports criminology and social harm, often employing online and offline ethnographic research methods.

Miria Grisot is Professor at the Department of Informatics, University of Oslo. Her research focuses on the sociotechnical study of digital infrastructures in healthcare.

Helene O. I. Gundhus is Professor at the Department of Criminology and Sociology of Law, University of Oslo, and Professor II at the Norwegian Police University College. Her research interests include policing and technology, police professionalism, migration control, crime prevention, and security.

Alexandra Hall is Associate Professor of Criminology at Northumbria University. She is a qualitative researcher with expertise in drug markets, organised crime, criminological ethnography, social and environmental harm, and the crime-economy nexus.

Keith Hayward is Professor of Criminology at the Faculty of Law, University of Copenhagen, Denmark. He has published widely in the areas of criminological theory, terrorism, popular culture, and the crime–technology nexus.

Thomas J. Holt is a Professor in the School of Criminal Justice at Michigan State University. His research focuses on various forms of cybercrime, cyberterrorism, and the policy response to these challenges.

Karen M. Holt is an Assistant Professor in the School of Criminal Justice at Michigan State University. Her research interests include sexual deviance, sexual offending, and the intersection of media and offending.

Julien Jeandesboz is Professor of European and International Politics at the Université libre de Bruxelles (ULB). He works on technology, security, border and immigration enforcement.

Mareile Kaufmann is a Professor at the University of Oslo, Department of Criminology and Sociology of Law. She works on the role of surveillance, science, and technology in different criminological arenas.

Emils Kilis is Senior Researcher at Baltic Studies Centre. His research interests include the social study of science, technology, and expertise in interdisciplinary and transdisciplinary contexts, and the impact of digitalization on various fields of human activity.

Susanne Krasmann is Professor of Sociology at the Department of Social Sciences, Universität Hamburg. She works on changing technologies of control and what they tell us about contemporary societies.

Malcolm Langford is a Professor at the University of Oslo, Department of Public and International Law. He works on human rights, public law and adjudication, new technologies, and legal education.

Matthias Leese is Assistant Professor at the Department of Humanities, Political and Social Sciences, ETH Zurich. His work investigates digital transformation and its effects on social and political ordering.

Naomi Lintvedt is a Doctoral research fellow at the University of Oslo, Norwegian Research Center for Computers and Law. Her research is on the regulation of privacy in human–robot interaction, under the aegis of the project Vulnerabilities in the Robot Society (VIROS) funded by the Norwegian Research Council.

Heidi Mork Lomell is a Professor at the University of Oslo, Department of Criminology and Sociology of Law. Her research interests include knowledge production, crime prevention, policing, and crime policy.

Rose Lunde is a PhD candidate at the University of Oslo, Department of Criminology and Sociology of Law. Her work focuses on prisons, punishment, and drug use treatment and recovery.

Jenny Maria Lundgaard is an Associate Professor at the Norwegian Police University College. She holds a PhD in criminology and does research on technologies and knowledge production in the police.

David Lyon is Professor Emeritus of Sociology and of Law, and former Director of the Surveillance Studies Centre at Queen's University in Kingston Ontario. His current research is on the medieval origins of modern surveillance.

Shoshana Magnet is Professor of Feminist and Gender Studies at the University of Ottawa.

Monique Mann is a Senior Lecturer at the Institute of Criminology at Victoria University of Wellington. Mann examines the political economy and social impacts of new technologies, relationships of power and influence, and advocates for enhanced protections to advance social justice.

Annette Markham is Chair Professor of future literacies and public engagement in the Department of Media and Culture Studies, Utrecht University, Netherlands, where she studies the lived experiences of disruptive technologies and ethical approaches to social research methods in digitally saturated contexts. Learn more at https://annettemarkham.com

Anne-Marie McAlinden is (full) Professor of Law and Criminal Justice, School of Law, Queen's University Belfast (UK). She is the author/editor of six books on sexual violence and abuse, including the dynamics of harmful sexual behaviors and cultural, legal, and justice responses.

Avital Mentovich is a Professor of Criminology at the Faculty of Law of the University of Haifa. Her research interests include procedural justice, legitimacy, and people's perceptions of legal institutions.

Sanja Milivojevic is Associate Professor in Digital Futures at Bristol Digital Futures Institute, the University of Bristol. She works on two key areas—borders and mobility, and crime and technology—in a range of criminological and interdisciplinary contexts.

Marion Oswald MBE is Professor in Law at Northumbria University, United Kingdom and Senior Research Associate at the Alan Turing Institute. She researches the interaction between law and digital technology and has a particular interest in the use of data and innovative technology by criminal justice bodies.

Elena Parmiggiani is Associate Professor at the University of Science and Technology (NTNU), Department of Computer Science, and senior researcher at Sintef Nord. She studies digital transformation of data management in digital platforms and infrastructures in different empirical domains.

Angela Paul is a Research Fellow in Law at Northumbria University, United Kingdom. Her research interests include public law, law enforcement technologies, and digital administrative processes.

Anastasia Powell is Professor of Family and Sexual Violence in Criminology & Justice at RMIT University, and a board director of Our Watch—Australia's national organization for the prevention of violence against women. She specializes in policy responses, legal reform, and primary prevention of interpersonal violence, including technology-facilitated abuse.

Orna Rabinovich-Einy is a Professor at the Faculty of Law of the University of Haifa. Her research interests include procedural law, dispute resolution, online dispute resolution, empirical analysis of the law, access to justice, and law and inequality.

Sofia Ranchordás is a Full Professor of Administrative Law at Tilburg Law and Professor of Public Law, Innovation, and Sustainability at Luiss Guido Carlo, Italy. Ranchordas works on regulatory studies, the digital transformation of the public sector, and is interested in how AI exacerbates socioeconomic inequities.

Helen Rand is a Senior Lecturer in Criminology at the University of Greenwich. She researches sex work with a particular focus on the roles of platforms in digital sex work.

Joanna Redden is Associate Professor at Western University. Joanna Redden researches the social and political implications of the increasing uses of AI and automated decision-making systems.

Kristin Bergtora Sandvik (Harvard Law School S.J.D 2008) is a Professor at the University of Oslo, Department of Criminology and Sociology of Law, and a Research Professor in Humanitarian Studies with the Peace Research Institute Oslo. Professor Sandvik teaches ethics, sociology of law, legal tech, and AI and robot regulations.

Peter Scharff Smith is a Professor at the University of Oslo, Department of Criminology and Sociology of Law. He works on punishment, prisons, human rights, sentiments on justice, the Holocaust, and war crimes.

Gregory Stratton is a Senior Lecturer in Criminology & Justice Studies at RMIT University. Greg also manages the *Bridge of Hope Innocence Initiative.* His research examines wrongful conviction, state crime, media and crime, and identity in the digital age.

Maryja Šupa is an Associate Professor at the Department of Criminology, Vilnius University, Lithuania. Her main interests are online offending research; algorithmic ethics and social control; and deviance in emerging technologies.

Amelia Svensson is an advisor and analyst on ID-misuse at the Norwegian National ID-Centre. She works on sociology of law, intelligence analysis on ID-misuse, crime prevention, and supporting and strengthening ID-work nationally and internationally.

Terrie Lynn Thompson is a Senior Lecturer in Digital Media and Professional Education at the University of Stirling. Her research examines the digital mediation of work and learning, and the innovative use of more-than-human theory and methods to understand the politics and ethics of what digital things do.

Nanna Bonde Thylstrup is Associate Professor at the Department of Arts and Cultural Studies, University of Copenhagen. Her research focuses on the politics and ethics of data, machine learning and digital infrastructures. She is particularly interested in how digitization and algorithmic processes are changing how we encounter, govern and practice knowledge infrastructures, and the political and ethical implications of these changes.

Daniel Trottier is an Associate Professor at the Department of Media and Communication at Erasmus University Rotterdam. His research focuses on forms of digitally-mediated visibility and social sanction.

Mirjam Twigt is a Research Officer/Postdoctoral Researcher at the Leiden Delft Erasmus Centre for Governance of Migration and Diversity, based at Leiden University, Institute of History. She works on the ways in which legal and digital infrastructures interact with people's experiences of unequal mobility.

Meropi Tzanetakis is an Assistant Professor in Digital Criminology at the University of Manchester. Meropi has been conducting pioneering research on the socio-technical aspects of digital drug markets and interdisciplinary studies on AI trust, security, and governance.

Pamela Ugwudike is Professor of Criminology at the University of Southampton, a Fellow of the UK's Higher Education Academy, and a Fellow of the national Alan Turing Institute for Data Science and Artificial Intelligence. She has international expertise in multidisciplinary research on the ethics and governance of Artificial Intelligence (AI) systems, with a focus on the data-driven technologies deployed by justice systems.

Rosamunde Van Brakel is Assistant Professor and postdoctoral researcher at the Vrije Universiteit Brussel, Faculty of Law & Criminology. She works on the impact and harms of surveillance and artificial intelligence in different criminological settings.

Kristin Veel is Associate Professor at the Department of Arts and Cultural Studies, University of Copenhagen. She works on the integration of digital technologies in our everyday lives and imaginations.

Maja Vestad is a PhD Research Fellow at the University of Oslo, Department of Criminology and Sociology of Law. Her current work explores the impact of changing technologies on forensic science, policing, and criminology.

Tone Linn Wærstad is a Professor at the University of Oslo, Department of Private Law. Her research focuses on human rights law, private international law, and consumer protection law in the family law area.

Professor **Sandra Walklate** is Eleanor Rathbone Chair of Sociology at the University of Liverpool who is recognized for her work in victimology especially in relation to policy responses to violence against women.

David S. Wall, PhD is Professor of Criminology in the Centre for Criminal Justice Studies, University of Leeds, UK where he conducts interdisciplinary research into the development of cybercrimes and its policing. He has published a wide range of books and articles on his research interests and has a sustained track record of interdisciplinary funded research in these areas and he has been a member of various governmental working groups on cybercrime.

Jacqueline Wernimont is Distinguished Chair of Digital Humanities and Social Engagement and an Associate Professor of Film and Media Studies at Dartmouth College in the United States. She is a digital media scholar who specializes in mathematical and computational technologies and whose books include *Numbered Lives: Life and Death in Quantum Media* and the co-edited *Bodies of Information: Intersectional Feminism and Digital Humanities* (with Elizabeth Losh).

Dean Wilson is Professor of Criminology in the Department of Sociology and Criminology, University of Sussex, UK. His research focuses upon surveillance, technology, policing, and security.

Dana Wilson-Kovacs is Associate Professor of Sociology at the University of Exeter, in the Department of Social and Political Studies, Philosophy and Anthropology. Her recent work is on the interface between forensic support and investigative practices and focuses on the use of digital forensics in policing in England and Wales.

Mark A. Wood is a Senior Lecturer in criminology at Deakin University. Most of his research examines the technology-harm nexus and he is currently completing a book on technology-harm relations that offers a framework for examining how technologies contribute to social harms.

Aleš Završnik is a Professor at the Faculty of Law at the University of Ljubljana (Slovenia) and a director of the Institute of Criminology at the same faculty. His research interest lay at the intersection of law, crime, technology, artificial intelligence, and fundamental rights.