SPRING 2019 / ISSUE 09

CREST SECURITY REVIEW

ISSN 2398-0540 (PRINT)
ISSN 2398-0559 (ONLINE)

Data

CREST SECURITY REVIEW
Editor – Matthew Francis
Guest Editor – Debi Ashenden
Illustrator – Becky Stevens
Designer – Steve Longdale
To contact *CREST Security Review* please email
csr@crestresearch.ac.uk

CONTENTS

Highlights

LEVERAGING LANGUAGE DATA

Computational Language Analysis can help profile the person behind a text and give us important clues about their future behaviours – p6

THE PROBLEM OF PREDICTING SOME OF THE PEOPLE, SOME OF THE TIME

While patterns of data exist, it is difficult to know which elements apply to subjects of interest. We still need to rely on humans to help us navigate ecological inference problems – p14

FROM THE EDITOR

This issue of *CREST Security Review* focuses on Data, and in particular on how the social and behavioural sciences can help us see the value that data and computer science can bring to understanding and countering security threats.

I'm grateful to our guest editor, Debi Ashenden, for pulling together this issue. She has drawn on some excellent research to highlight the challenges in managing data as well as showcasing the potential benefits for how we use data to make better decisions faster.

Debi has provided a helpful overview of the issues (page 4), and of some of the topics covered in this issue's articles. She has also drawn on one of her ongoing projects to give us an insight into how algorithmic decision making can be improved, to help build trust between governments and citizens (page 12).

On the theme of data, Paul Taylor (page 14) focuses on the challenges of applying big data solutions to small data problems. He shows us how we might approach problems like where we are able to identify patterns in terrorists' behaviour, but still unable to build a predictive terrorist 'profile'. Jo Hinds on page ten also addresses risks and opportunities in using big data to predict behaviour.

Turning away from the challenges in this area, on page six Ryan Boyd and Paul Kapoor look at how Computational Language Analysis can help profile the person behind a text and give us important clues about their future behaviours. Providing a cautionary note to the use of language, Pip Thornton highlights how linguistic data is mediated and manipulated by large technology companies (page 8).

Last, or perhaps most usefully first, Duncan Hodges (page 24) provides us with an A-Z of key terms in discussions of data – you'll never again be found wanting if the discussion turns to Zipf's Law.

We have two articles in this issue looking at problems encountered in eliciting information. On page sixteen, Feni Kontogianni presents her research into the problems in recalling detail from repeated events, and presents some of the techniques for overcoming them. On page eighteen, Pamela Hanway looks at how we can reduce the cognitive load on interviewers.

Joel Busher, Donald Holbrook and Graham Macklin summarise the findings from their recent project that helps explain why some extremists or groups choose not to engage in violence (page 22). Looking at a different aspect of problems with extremism, Sarah Marsden (page 20) provides us with an introduction to good practice in countering violent extremism, based on a recent guide she's released on this subject.

This is our first issue to go out simultaneously in print, online and in our mobile app. We created the mobile app to help widen the audience of people who can access the research we feature. Like the magazine, we've aimed for an experience that presents rigorous science in an accessible way that looks good. We value your feedback on how you find using the new portals (see my email address below).

The new app is accompanied by a new website to host *CSR*. You can find it at www.crestsecurityreview.com. You can also find details on how to download the app on the back cover. Please do tell your colleagues and friends.

Finally, don't forget that you can read more about some of the research featured in this issue in our Read More section on page twenty-six, and please get in touch if you have ideas of research that you'd like to see featured in future issues. You can email these to me at m.d.francis@lancaster.ac.uk

Matthew Francis
Editor, *CSR*

DEBI ASHENDEN

DATA AND THE SOCIAL AND BEHAVIOURAL SCIENCES

The science of how we manage and leverage data is unsurprisingly an increasingly ubiquitous topic. Data is often thought of as the base of a pyramid on which information, knowledge and wisdom sit. Data science is the extraction of information from data with the aim of developing knowledge.

The emergence of data science is driven by our aspirations to make better decisions faster through automation by leveraging the unprecedented amounts of data that we can now gather, as well as exploiting our ability to design algorithms and take advantage of increased computing power.

The impact of advances in data science has the ability to touch all aspects of daily life from the so-called 'datafication' of society through to the 'quantified' self. Automated decision making is providing benefits in financial transactions, the delivery of personalised services online, health care prediction and diagnosis, and the development of government services.

However, automated decision making also has the potential to discriminate against individuals leading to the denial of some services. Further, the lack of transparency in algorithm design and implementation can cause distrust and potential social unrest. Advances in data science are not confined to social applications, the exploitation of data is unsurprisingly of interest to defence and security practitioners.

In a public speech at St Andrew's University, the Director of the UK's Secret Intelligence Service, Alex Younger, highlighted the importance of achieving mastery in the data age. He also talked about the changing context where adversaries do not see a clear delineation between war and peace.

The UK's Ministry of Defence has a similar focus on the better use of data and has issued a Joint Concept Note on Information Advantage (JCN 2/18), highlighting the way that adversaries are using advances in technology to achieve '*mass customisation of messaging, narrative and persuasion*' that extends both reach and influence. Actions by adversaries often now take place in the 'grey zone' between war and peace, frequently targeting broader society with the aim of creating uncertainty, ambiguity, doubt and undermining confidence in decision making.

It is clear that defence and security practitioners need to be able to balance taking advantage of data with ensuring that decision making processes are resilient to both attack and to misuse. As Russia expert Keir Giles has pointed out, this means understanding both the content of information processes as well as the code that underpins them.

The need for understanding spans the requirements for an algorithm, the theory that underpins the design, and construction as well as training in how data is used.

In this issue of *CREST Security Review,* we see the value that data and computer science can bring to topics such as computational language analysis for understanding the person behind the text (Ryan Boyd and Paul Kapoor). The article by Joanne Hinds highlights the potential benefits of data for predicting behaviour, while sounding a note of caution around ethical issues.

The article by Pip Thornton continues this theme by pointing to the impact that digital capitalism can have on spreading fake news, while my article on algorithmic decision making highlights the impact that conceptual models that underpin automated decision making can have on the relationship between individuals and the state. Fortunately, there are research institutes set up with the aim of addressing some of these issues. In the UK the **Alan Turing Institute** for data science and AI is well established and has a defence and security research theme within its programme.

The focus of the institute, however, is on the key disciplines of mathematics, engineering and computing. While these are of vital importance for the development of data science, algorithms are ultimately deployed in a real-world context. The aim of the Institute is to 'change the world for the better', but it is incumbent on researchers to critique this statement – who constitutes 'the world' in this instance and 'better' for whom? Fortunately, the recent establishment of the Ada Lovelace Institute (and the close working relationship between the two) provides balance. The **Ada Lovelace Institute** has the aim of ensuring that, 'data and AI work for people and society' and considers the impact of data science on society.

There are many other research initiatives that are at different stages of maturity and which address some of the emerging issues of data and data science. For example, the **Data Justice Lab** recognises that if data is misused it can heighten socio-economic inequalities and has the potential to increase social divisions. The **Not Equal Project** focuses on the socio-technical aspects of new technology considering how it can empower, emancipate and offer opportunities for economic development.

The **Unbias Project** considers ways of improving algorithmic transparency to build trustworthiness in systems. The **People Powered Algorithms for Desirable Social Outcomes** project looks at the design of algorithms and aims to understand how algorithms mediate real world relationships between the state and individuals.

Research questions around data science topics in general and automated decision making more specifically are still emerging in the defence and security space, not least because the focus of data science is on developing automated decision making processes through Artificial Intelligence and Machine Learning (AI/ML), whereas decision making is an inherently human activity. Users of automated decision making tools may feel reluctant to rely on an algorithm so how do we understand how to build trust in algorithms? Is it more acceptable for a human to make a poor decision than it is for a machine? Do we expect more from automated decision making than it can truly deliver at the moment? How do we protect algorithms during the design and development phase to ensure that training data, or the algorithms themselves are not tampered with? How do we ensure that algorithms are designed on robust theoretical principles – that they are actually doing what we want them to do? This issue of *CREST Security Review* starts to explore the topic but it is evident that there is still much that social and behavioural science can contribute to ensuring that the aspirations of data science are met for defence and security practitioners.

Professor Debi Ashenden is the guest editor for this issue of CREST Security Review. *She is Professor of Cyber Security at the University of Portsmouth (UK), Research Professor of Cyber Security and Human Behaviour at Deakin University (Australia) and leads CREST's Protective Security and Risk programme.*

RYAN L. BOYD AND PAUL KAPOOR

PSYCHOLOGICAL PROFILING AND EVENT FORECASTING USING COMPUTATIONAL LANGUAGE ANALYSIS

Psychologists have long believed that we can discern what makes a person tick by analysing their language. The modern study of language has become a highly sophisticated area of research that leverages computational modelling, objective measures of language, and extensive empirical rigor.

The links between a person's mental processes and the words that they say or write have been extensively studied, validated, and applied to fields as diverse as computer science, medicine, sociology, and anthropology, to name just a few. The ability to 'get inside a person's head' by analysing their language patterns from a distance has tremendous appeal and several practical applications, ranging from the patently obvious to the surprisingly nuanced.

SUBSTANCE VERSUS STYLE

In research on the psychology of language, most scientists have traditionally focused on the substance or content of language – words that have an explicit meaning (e.g., house, friend, bomb, etc.). Unsurprisingly, there are several direct links between what a person talks about and what they are thinking. Extroverts tend to use more words related to social processes and use more positive language. Neurotic people tend to use more words indicative of anxiety, and so on.

These (perhaps obvious) patterns are grounded in psychological theory and can therefore be extrapolated to a broader understanding of the individual. The content of a person's language is reliably diagnostic of their intelligence, political orientation, personality characteristics, and even how long they live. The things that occupy a person's mind are not merely diagnostic of their thoughts – they are indicative of deeply ingrained patterns in their life.

Perhaps more interesting, however, is the style of a person's language. An abundance of research in recent years has found that the small, throw-away words in language like articles, pronouns, conjunctions, and so on, are deeply revealing of lower-level psychological processes. People whose language is highly self-focused (e.g., high rates of words like 'I', 'me', and 'my') tend to be relatively insecure and depressed. People who use more articles ('the', 'a', 'an') and prepositions ('in', 'by', 'across') in their writing tend to be more analytic in their thoughts, and factors such as someone's social status and authenticity tend to be reflected in a person's linguistic style more than its content.

By pairing the analysis of what a person says with how they say it, we can often paint a remarkably detailed picture of a person's mental and social universe. Such analyses can be performed extremely quickly and objectively using computational tools, and many psychological phenomena can be reliably estimated using relatively simple statistical models.

PSYCHOLOGICAL PROFILING

Much of the work in computational psychological profiling is founded on research demonstrating that linguistic patterns are relatively stable across time and contexts, particularly the stylistic components of language. The quality of language-derived psychological profiles can range from speculative to unbelievably strong, sometimes allowing us to identify an author with near-perfect accuracy using only their language. A language-driven approach to profiling allows us to understand the person behind a given text rather than just the text itself.

Rather than simply taking a threat of violence at face value, we can computationally evaluate the speaker's language for deeper clues. Are they at-risk for a future depressive or schizophrenic episode? Are they obsessive-compulsive, or perhaps prone to conspiratorial thinking? Statistical estimation of these types of psychological vulnerabilities can help to highlight critical intervention strategies.

Language-based psychological profiles can also be applied at the group level, revealing fundamental differences in how group members think and engage with the world. Recent research found that Islamic State, as a group, shows greater authoritarianism and religious fervour in their psychological profile (revealed by markers such as low rates of present-focused and tentative language, plus high rates of religious language) relative to al-Qaeda. Moreover, study participants who scored high on authoritarianism and religiosity reported more favourable attitudes towards the language of Islamic State compared to the language of al-Qaeda. Understanding such group differences can provide insights into how a group functions, as well as what types of people might find these groups appealing.

Psychological profiles can also be built for broader communities and monitored over time. The psychological health of a community can easily be tracked following a tragedy using various data sources, such as newspapers or social media. Research from several labs has investigated the psychological impact of calamitous events ranging from the 9/11 attacks to mass shootings, finding unique patterns of coping as they unfold in response to major upheavals.

BEHAVIOURAL FORECASTING

In a vacuum of information about an author, the statistical analysis of language can give us important clues about a person's future behaviours. Often, this approach relies on the relationship between language and general behavioural patterns. For instance, we find that the language representative of someone's core values (e.g., family, work ethic, empathy) are strongly related to their regular behaviours, such as attending religious functions, donating time/money to a cause, or even playing games online.

Most of the recent work in language-based behavioural forecasting focuses on interpersonal behaviours. Messages written prior to events like suicide or spree killings show distinct psycholinguistic fingerprints. When soliciting sex from minors online in sting operations, individuals who exhibit high certainty and planning markers in their language are at high risk for repeat-offending in the same crime categories (e.g., acquisition of child pornography, future attempts to solicit minors).

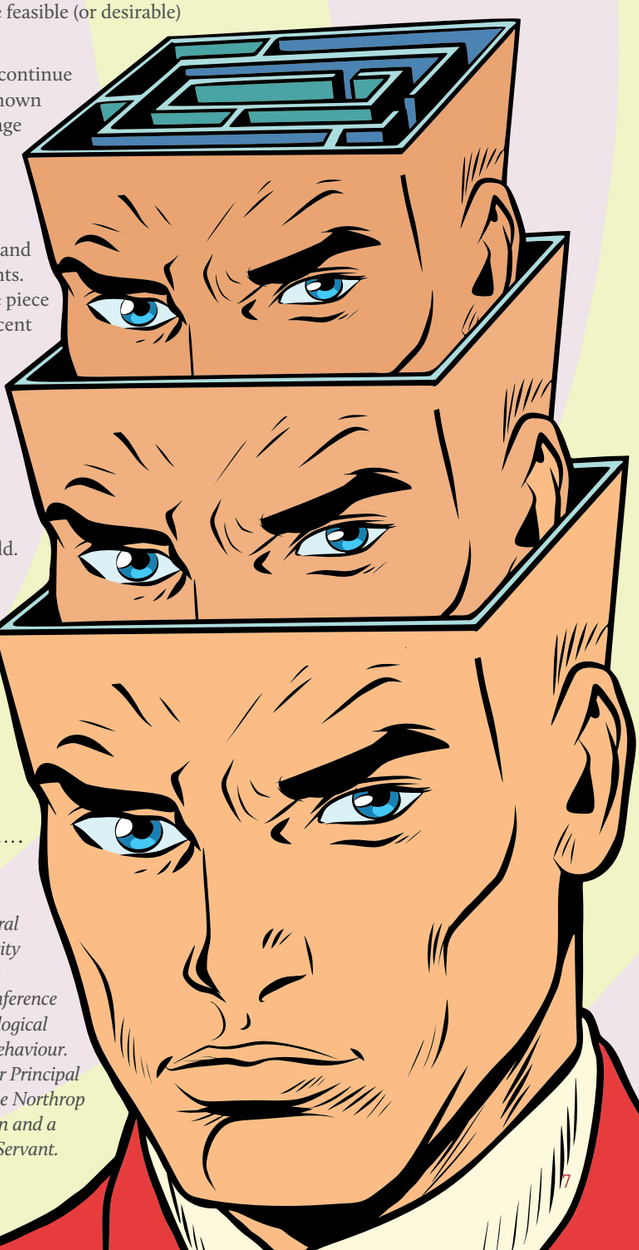
Similarly, research on group processes finds that linguistic cues related to planning decrease immediately before the betrayal of an ally (along with an increase in positivity and politeness). A failure to linguistically adapt to a changing group membership tends to precede members exiting a group, and changes in interpersonal linguistic coordination can foreshadow the initiation, stability, and dissolution of a relationship.

WHERE DO WE GO FROM HERE?

The implications of language-based profiling and behavioural forecasting are far-reaching and can represent a double-edged sword. The same language data can be leveraged for multiple purposes, and care must be taken to protect the words of vulnerable or high-profile individuals. Individuals working in the security sector who are psychologically vulnerable can be identified from their language patterns, resulting in a non-negligible risk for targeted exploitation.

Sources who have insight into future plans require particularly high discretion. A person who knows of impending policy changes or upcoming events may show extremely subtle changes in their language patterns. Such changes are often not discernible to an untrained observer yet can still be detected using modern computational techniques. Compartmentalisation of knowledge may serve as insulation from detection, yet this approach may not be feasible (or desirable) in many situations.

Future research will continue to discover still-unknown links between language and psychology, meaning that language data from any period can be revisited extensively and mined for new insights. Language data is one piece of the puzzle, and recent work that integrates language-derived profiles with other known factors (such as age, political affiliation, and images) show significant promise for advancing the field. Reliable obfuscation techniques remain to be developed and will likely be reactive (rather than proactive) as new methods for language analysis continue to emerge.



Dr Ryan L. Boyd is a computational social scientist and behavioural scientist at the University of Texas at Austin. His research involves the inference of motives and psychological patterns from verbal behaviour. Paul Kapoor is a Senior Principal Systems Engineer at the Northrop Grumman Corporation and a former US Navy Civil Servant.

PIP THORNTON

WORDS AS DATA: THE VULNERABILITY OF LANGUAGE IN AN AGE OF DIGITAL CAPITALISM

The security of the data that circulates the internet is dependent on much more than cryptographic key exchange. Data can represent all manner of information that might threaten personal and national securities and safety, be it through the misuse of social media or mapping data, the tracking of personal information for advertising, or the state-led gathering of financial or communications data.

Some of these data (mis)uses can of course be avoided, mitigated or challenged, but there is one type of data that underpins almost every aspect of our digital lives, regardless of who we are, which is much harder to shield from forces of commercialisation, surveillance and the systemic biases of technology – and that is linguistic data.

The language that flows through the platforms and portals of the Web is increasingly mediated and manipulated by large technology companies that dominate the internet, and in particular for the purpose of advertising by companies such as Google and Facebook. Whether through keyword targeting, email, search engine optimisation techniques, or the dissemination of news or status updates, the words that circulate through digital space are increasingly laden with economic value.

In this respect, words-as-data become detached from their original function as a means of human communication, and instead become vessels for the flow of advertising and cultural capital around the online and offline world. This has significant consequences.

We all need to communicate, access information and keep up in the modern marketplace, but in today's digitally networked society, the words we enter into Web-based platforms such as search engines and social media have themselves been turned into valuable pieces of data. And when words are digitised for transmission and processing through the Web, they lose their original context. Just like any other type of data, linguistic data becomes vulnerable to manipulation and monetisation.

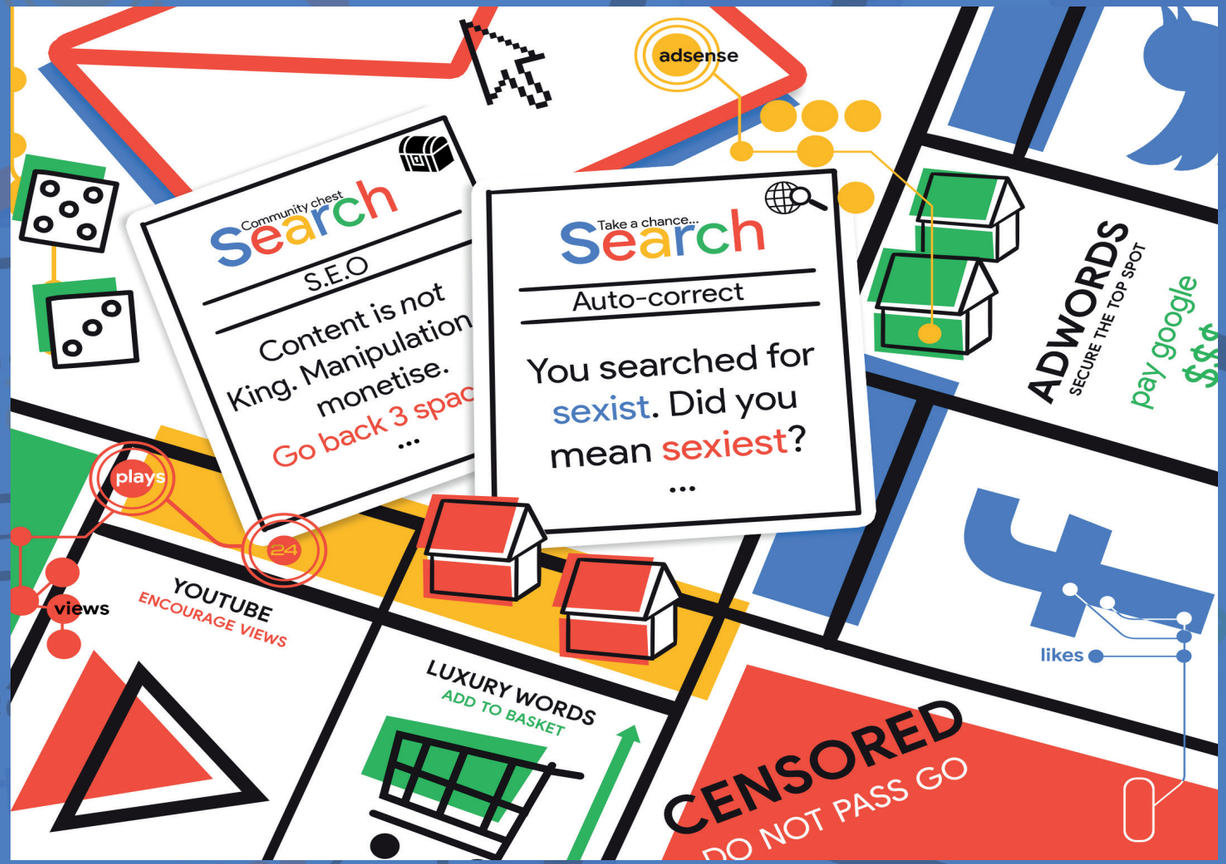
The computational manner in which linguistic data is processed is responsible for the sometimes amusing, but also sometimes dangerously stereotypical and controversial auto-predictions that appear when you start typing in the Google search bar. Auto-predictions are based on a mixture of aggregated previous searches, and the existing data available on the Web. Words and phrases that appear more frequently next to each other in this 'searchable database' will therefore be more likely to complete your search query.

The problem with this is that any omission, manipulation or bias in the searchable database is therefore reproduced and compounded. So the word 'man' or 'male' might be more often associated with nouns like 'doctor', 'boss' or 'CEO', and this will be reflected in search results and auto-completions. It is also the reason why online translation services like Google Translate are often so bad.

Google can at any time also interfere, censoring certain keywords so that they won't be included in the construction of search results. This might be for political, commercial, legal or ethical reasons. Google is not a neutral and democratic gatekeeper of the world's information, and it is crucially important not to treat what comes out of the search engine as unmediated truth.

The way digitised language is structured is also dependant on the monetary value of words in the online advertising industry. Google is one of the main players in this marketplace – its commodification and exploitation of language has been described as a form of 'linguistic capitalism'. Google has around a 95% market share of internet searches in the UK, and its advertising platforms AdWords and AdSense have an ever increasingly significant impact on how all kinds of information circulates on the Web.

AdWords is the system by which advertisers bid and pay for keywords and phrases in order to secure the top spots on Google's search engine results page. Each time somebody searches for a word on Google, a mini auction takes place, and the advertiser with the highest bid for that particular word at that time wins, and as long as their advert is considered worthy by the algorithmic ranking system, their advert will appear at the top of the search results page, above the 'organic', non-paid results. The information appearing before our eyes is therefore mediated by the vagaries and complexities of a linguistic market. Even the so-called 'organic' search results are significantly affected by the forces of linguistic capitalism. A whole Search Engine Optimisation industry has grown out of identifying and valuing keywords to make online text more attractive to search algorithms.



And this is a really important point. Much of the text that exists online is structured and restricted by digital processing systems, and/or created or optimised not for human readers, but for the algorithms that scrape text for the purposes of targeted advertising. The information we receive through search engines is therefore susceptible and vulnerable to the fluctuations and restrictions of an algorithmic marketplace. The value – and therefore the reliability – of language has become destabilised by digital capitalism.

Digital capitalism also has a huge role to play in the rise of fake news. While propaganda and subversive advertising are nothing new, many of the 'fake news' stories that circulated the Web in the run up to the 2016 US Presidential election were written not for any particular political motive, but because Google pays website owners to host adverts through its AdSense platform. The more views a website (and the adverts served on it) has, the more money the owner makes, regardless of its content.

A politically controversial story, spread virally through media such as Facebook 'likes', 'shares' and 'comments', can generate thousands of dollars in advertising revenue. What is important to remember here, is that the stories being generated, while often completely made up – as in the case of many of the anti-Clinton stories in 2016 – become embedded into the fabric of the Web, their linguistic data contributing to future searches, translations, and other informational systems.

The influence and control of language on the Web therefore translates into a frightening power over the generation and dissemination of information. As a result, we need to be asking what narratives are we creating when our online discourse is optimised for the spread of capital rather than for narrative communication? What does it mean that every query we make of a search engine is influenced by (often opaque) algorithmic 'market forces', or that YouTube videos aimed at children contain sexual or violent material to encourage more views and therefore more advertising revenue? As we have seen in the revelations about Cambridge Analytica, the spread of fake news through digital advertising is perhaps the tip of the iceberg.

The systemic manipulation and monetisation of digitised language is a threat to the security and stability of modern society. The very words we use to communicate, learn, debate, and critique have become compromised by opaque algorithmic organisation and optimisation, and the market-driven profits of private companies such as Google. We might therefore ask ourselves, just how resilient and secure is language in the digital age? Indeed, how can we even talk about security when we cannot talk securely?

Pip Thornton is a Post-Doctoral Research Associate in Creative Informatics at the University of Edinburgh. The material in this essay is based on her recently published articles 'A Critique of Linguistic Capitalism: Provocation/Intervention' *(2018) and* 'Geographies of (con)text: Language and Structure in a Digital Age' *(2017), and on her research blog www.linguisticgeographies.com*

JOANNE HINDS

BEHAVIOUR PREDICTION: THE CHALLENGES AND OPPORTUNITIES OF BIG DATA

We base many of our decisions about other people on assessments such as what we think their personalities are like, or how they may behave in certain situations. Our ability to judge others accurately can have profound consequences in terms of who we socialise with, date or employ.

In physical environments, we use 'cues' such as a person's voice, dress or demeanour to form our judgments. Online, we may use their Facebook profiles, blogs or tweets. The online equivalent of these cues are often referred to as 'digital footprints' or 'digital traces'. These provide opportunities to analyse individuals' attributes and behaviour at mass scale and over long periods of time. So, as our interactions with technology continue to increase, can data be used to infer who we are and how we might behave?

PREDICTING BEHAVIOUR

Using data to predict behaviour has many applications including healthcare, marketing, and criminal investigation. In recent years, academics within psychology and computer science have examined the extent to which individuals' information can be inferred from their digital data. In particular, researchers have attempted to predict individuals' personality traits and demographic attributes.

Personality traits are emotions and behaviours that make up an individual's idiosyncratic disposition. The 'Big Five' (also known as the Five-Factor or OCEAN model) is the most popular approach currently used by researchers when measuring personality. Assessments consist of self-report questionnaires, which evaluate how highly individuals score across five dimensions as follows:

OPENNESS Have a variety of interests/hobbies, enjoy travel/adventure and are comfortable with change.

CONSCIENTIOUSNESS Highly organised, possess leadership skills, prefer planned activity over spontaneous behaviour.

EXTRAVERSION Sociable with many friends, outgoing and talkative, likely to participate in sports.

AGREEABLENESS Highly compliant, forgiving, cooperative and may be perceived as being a pushover.

NEUROTICISM Prone to depression, anxiety, low self-esteem as well as general negative emotions toward situations.

Demographic attributes can relate to any aspect concerning an individual's background characteristics or socioeconomic status. Predicting individuals' demographic attributes is well established in areas such as computer forensics and computational linguistics which often use text-based sources to predict an individual's age and gender. More recently, researchers have used digital data to predict other attributes such as location, occupation, level of education, sexual orientation, and political preferences.

Studies have also used digital data to successfully predict election results and reactions or opinions to events such as the Arab Spring and even box office revenue for films. For example, in the latter case, Márton Mestyán and colleagues demonstrated that the popularity of a movie could be predicted by editor and viewer activity on the film's Wikipedia entry.

HOW BEHAVIOUR CAN BE PREDICTED ONLINE

Similar to the way in which humans use cues to formulate opinions of other people, computer algorithms use 'features', where digital traces (e.g., Facebook likes, number of followers) are analysed to establish the strength to which they are associated with particular attributes (e.g., age, extraversion, location). Typically, this 'experiment' is performed on a subset of data, and then this subset is used to 'train' an algorithm to predict said attributes from the remainder of the dataset. The accuracy of the algorithm's performance then informs the researchers how successful their prediction was.

The ability to predict individuals' personal information, preferences and behaviour can have welcome effects and positive outcomes. Recommender systems such as Netflix or Amazon provide users with suggestions of films and products that individuals are likely to enjoy based on their previous activity.

Likewise, targeted marketing derived from our previous behaviour can be useful when individuals are exposed to advertisements that align with their personal preferences. However, such approaches are far from perfect and can sometimes be inconvenient or annoying. For instance, users who have purchased household items on Amazon go on to receive countless ads for the same items months later.

PREDICTING BEHAVIOUR: ETHICS AND CHALLENGES

Unfortunately, predicting information from digital data can extend beyond mere irritation to unintended or malicious consequences. For instance, individuals who are similar (in age, location, interests etc.) tend to be friends with, or connected with each other. Indeed, the notion that birds of a feather flock together (also known as 'homophily') is a truism often reflected in individuals' online social networks.

These patterns in online human networks can therefore create the potential for shadow profiling – where an individual's undisclosed or private information is revealed or inferred from data accessed through other people within their network. Recent research has emphasised the dangers of shadow profiling by demonstrating the potential to infer the sexuality of non-users of social networking sites.

The potential for shadow profiling highlights just one example of the type of ethical challenges surrounding the privacy and security of peoples' data.

The introduction of the EU General Data Protection Regulation (GDPR) that came in to force in May 2018, attempts to address more recent issues in terms of how personal data is handled. And whilst more up-to-date regulations are certainly necessary and beneficial, it is incredibly difficult to know the true extent to which technology will impact our lives (and our data) in the coming years.

The scale of the challenge is demonstrated by current estimates that predict around 30 billion online devices will be connected to each other by 2020. At the same time as these devices are generating data, data breaches occurring across banking, healthcare and technology companies (e.g., the WannaCry ransomware) have demonstrated the widespread threats to people's data across numerous industries.

In the case of the Cambridge Analytica scandal, data from approximately 87 million individuals' Facebook accounts were collected without their explicit consent.

Data like these were supposedly used to create targeted advertisements, such as those which attempted to influence people's voting preferences in the 'Vote Leave' campaign in Britain's European Referendum, and Donald Trump's 2016 presidential election.

Amidst concerns of how data are collected, used, shared and what true 'informed consent' really is, many people feel uncomfortable with the notion that their devices are 'listening' or that their behaviour is constantly being monitored or analysed. Whilst it is an exciting time for technological advancement and social science, organisations and cybersecurity practitioners face some complex challenges when it comes to handling data carefully and reinforcing trust in using technology.

Dr Joanne Hinds is a Research Associate at the University of Bath. Her work focuses on predicting information and behaviour from digital traces using psychological and computational techniques.

DEBI ASHENDEN

ALGORITHMIC DECISION MAKING

How can re-designing system interactions help build trust between governments and citizens, enhance the security and wellbeing of individuals and protect the security of the state? Debi Ashenden draws on research on people-powered algorithms to show some of the difficulties and solutions.

Public services are increasingly being delivered using data-driven decision-making algorithms. It is well understood, however, that our ability to develop data-driven solutions through the use of machine learning or AI, is currently outstripping our understanding of how to incorporate social norms in the technology being developed.

The result of this is that data-driven decision-making algorithms may offer efficient and effective ways of allocating resources to individuals, but the decisions made are often not seen by individuals as being legitimate or fair, leading to a mistrust in the system and a willingness to find ways to work round it.

The term 'algorithm' is invoked in different academic disciplines and across governments in a variety of ways. Given these differences, how can we be certain policy makers and data scientists discussing the requirement for an algorithm to automate public service delivery are talking about the same thing?

Without a dialogue between these two communities, how can we be certain that what is delivered by the algorithm is what the policy maker intended? If we're not clear about how social policy is abstracted into conceptual, logical and physical models, how can we have confidence in the resulting algorithm, or family of algorithms? We need to be able to reflect on and critique the assumptions that are currently being made in the design of automated decision making systems. By doing this we can start to have a discussion around how to address issues of legitimacy and fairness.

The introduction of a new UK system to manage social-welfare payments, Universal Credit, provides an example of the real-world problems we may well see with automated decision making for public services. The conceptual model for Universal Credit is based on conditionality, whereby social exchanges are defined in which the individual must participate in order to be in receipt of welfare.

Fundamentally, it is the social exchange itself that is subject to tests of legitimacy and fairness. The data that are input to the Universal Credit system determines the level of benefits to which the claimant is entitled, as well as under what conditions those benefits will be paid. This is a conditionality approach and to be seen as fair and legitimate by the user it requires predictability; it also implies reciprocity as both the state and the individual have to give in order to receive. The state offers protection to its citizen and in return they have to give up some freedoms.

In the digital environment, if the state fails to provide protection then in turn citizens may question why they have to give up any freedoms. In the example of Universal Credit, reciprocity becomes problematic when incorrect calculations are made; benefits may be either underpaid, or overpaid and then need to be paid back. Such problems expose the gap between what the state and the individual define as fair and legitimate but there is limited opportunity to discuss and negotiate the decision. Any discussion that does take place is between individual claimants who share their stories with each other and advise on potential solutions.

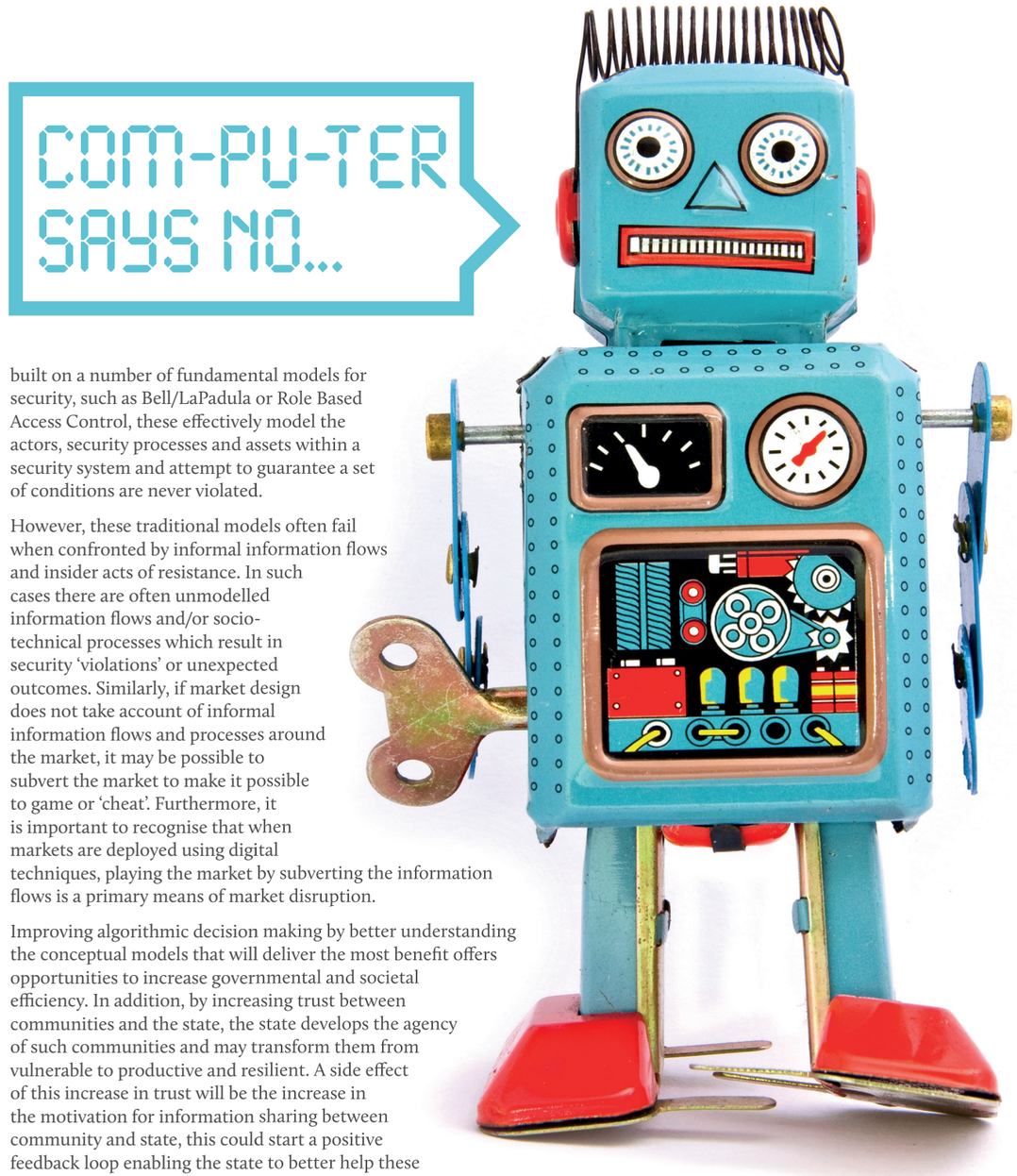
While it is necessary to model information flows, there may also be other types of decision-making techniques that could be used. These could provide more effective conceptual models for algorithm design, offering a potentially more secure and trustworthy system. Market design techniques are used in any scenario where one is trying to change features of social interactions involving a scarce resource in order to bring about an optimal solution for all parties.

Market design has been used in many areas where vulnerable people come into contact with governments under feelings of insecurity and alienation. Market designers have had dramatic success over the last thirty years promoting reform in areas as diverse as live organ donation, refugee resettlement, the allocation of children to schools, doctors to hospitals, public housing systems and the management of food banks.

Market designers have been closely led by empirical research which has led to clear policy prescriptions. In turn these have led to dramatic results. For example, the reform of American food banks led by Candice Prendergast increased the supply of food across the USA by roughly $100 million around the time of its introduction, purely through creating a more efficient and responsive system.

Market design offers an alternative conceptual model to the conditionality model in Universal Credit. Market design has been explored in relation to the resettlement of those fleeing conflict or persecution so that they can become productive and valued members of a secure society. Successful resettlement means that refugees are free from harm but also free to build new lives, both important elements of personal security.

However, there are subtler similarities in the manner in which both market designers and security architects model and explore their respective domains. Security systems are traditionally



built on a number of fundamental models for security, such as Bell/LaPadula or Role Based Access Control, these effectively model the actors, security processes and assets within a security system and attempt to guarantee a set of conditions are never violated.

However, these traditional models often fail when confronted by informal information flows and insider acts of resistance. In such cases there are often unmodelled information flows and/or socio-technical processes which result in security 'violations' or unexpected outcomes. Similarly, if market design does not take account of informal information flows and processes around the market, it may be possible to subvert the market to make it possible to game or 'cheat'. Furthermore, it is important to recognise that when markets are deployed using digital techniques, playing the market by subverting the information flows is a primary means of market disruption.

Improving algorithmic decision making by better understanding the conceptual models that will deliver the most benefit offers opportunities to increase governmental and societal efficiency. In addition, by increasing trust between communities and the state, the state develops the agency of such communities and may transform them from vulnerable to productive and resilient. A side effect of this increase in trust will be the increase in the motivation for information sharing between community and state, this could start a positive feedback loop enabling the state to better help these communities and carry out policy.

Addressing such questions will support the development of advances in data science by providing criteria that can be used to create conceptual foundations for automated decision making systems that are perceived as fair, inclusive and empowering. These advances in the understanding of the social acceptance of automated decision making will help to keep pace with the advances in the UK data science community. This will ensure the UK continues to be able to take full advantage of the research in data science, machine learning and AI.

Professor Debi Ashenden is Professor of Cyber Security at the University of Portsmouth (UK), Research Professor of Cyber Security and Human Behaviour at Deakin University (Australia) and leads CREST's Protective Security and Risk programme. People-Powered Algorithms for Desirable Social Outcomes is a collaboration with Professor Lizzie Coles-Kemp (Royal Holloway, University of London), Dr Oliver Buckley (University of East Anglia), Dr Duncan Hodges (Cranfield University) and Dr Will Jones (Royal Holloway). It is funded by EPSRC (EP/R033382/1).

PAUL TAYLOR

FROM DATA TO DATUM: WHAT SHOULD I DO IN THIS CASE?

At the heart of many scientific efforts to help security professionals is a mathematical challenge. One that has occupied the minds of biologists, sociologists, psychologists, and statisticians for decades. One that highlights both the power of cases and the limits of data. One that has no easy solution, though many have tried.

Its formal name is the ecological inference problem: the problem of making inferences about an individual from aggregate data (and data models). It is why evidence suggests that there is simultaneously no single terrorist 'profile' and yet common patterns across many terrorists' lives. Patterns exist, but it is difficult to know which elements of those patterns are relevant to 'that' individual.

What is observed 'on average' speaks to what is true for only some. Fortunately, our uncanny ability to make sense out of this noise in this 'social data' means that investigators are able to provide the necessary, nuanced perspective.

The consequence of being able to 'predict some of the people some of the time,' as a series of social psychology papers described it in the 1980s, depends on the investigative task. Data work well if you want to predict data. If your goal is to prioritise who to investigate or how much resource to allocate, then a statistical model that weights risk factors is likely better (and more ethical) than random investigation or random resource allocation.

Data is less powerful when you want to predict datum. The challenge was nicely illustrated in a recent review my colleagues and I undertook of deception detection methods. Typically social science in this area administers a technique, such as asking unanticipated questions or providing a 'model' statement for the interviewee to emulate, to one group of interviewees. Then they administer a standard questioning technique to a comparison group. The researchers then compare the two interviewee groups to determine whether, on average, the new technique elicited more information than the standard technique. In the case of unanticipated questions and model statements, they do. On average, people give up more information and their deception is better detected when these techniques are used in the interview, compared to when they are not used.

But, what about a single case? When using these techniques, what criterion – what count of the details the interviewee provides – should I use to infer that you are lying?

Our review found wide variation in the criterion that worked best for each study. So much so that any single criterion would result in us exonerating all liars, or falsely accusing all truth-tellers, in at least one study. The reason? Context is everything. What you're describing, what you've experienced, and how well you are interviewed, influences what you report far more than whether you are lying or telling the truth.

In the deception detection world, a recognition of the ecological inference problem has led researchers to focus on information elicitation, recognising that the only way to determine veracity for sure is to elicit a checkable fact, which can be verified elsewhere.

In other domains, the solution is coming from the coalescing of three efforts. The first is the development of more precise inference models. In the deception field, baselining an individual's behaviour or using a criterion that is culturally-specific improves the accuracy of predictions. In work predicting risk of violence, layering contextual moderators into an assessment of individual-level push and pull factors tends to provide a more nuanced view of how that factor should be weighted for that individual. The difficulty of this approach is that models quickly get complex and data too sparse for meaningful development and ethically-defensible use.

The second is the use of innovative, more discriminatory indicators. This is one area where social science is uniquely positioned to contribute. Theory-driven models can inform what new data we look for. While the development of precise inference models looks to squeeze the most out of existing data, this solution encourages us to be informed consumers. Less is often more. In all of the many models proposed for insider threat detection, often using hundreds of variables, one that has survived the most rigorous testing involves a single measure carefully derived to capture a unique aspect of insiders' experience — the inability or unwillingness to maintain normal interpersonal behaviour with co-workers.

The third is to recognise how good humans can be at navigating ecological inferences. Despite all our unconscious biases, we are uniquely disposed to infer the 'story' that underlies data and to form hypotheses that allow us to test such stories. We're good at finding ways to determine if the inference is correct on these occasions, so long as we receive feedback on the accuracy of our judgements over time. One interesting way to encourage this positive aspect of human inferences is to provide investigators a systematic way to capture and compare the assessments among colleagues. One CREST-funded project, led by Professor Ashraf Labib, is researching precisely this – see https://crestresearch.ac.uk/projects/taking-decisions-information-value/. Security and intelligence agencies will depend on their case officers even more because of, not in spite of, the increasing use of data.

Paul Taylor is the Director of CREST, Professor of Psychology at Lancaster University in the UK, and Professor of Human Interaction at the University of Twente in the Netherlands.

FENI KONTOGIANNI

UNTANGLING THE PAST: REMEMBERING DETAILS OF REPEATED EVENTS

If you only attended one meeting of a terror cell, it might be easier to remember who said what than if you attended several. Feni Kontogianni draws on her research to explain the problems in recalling detail from repeated events, and some of the techniques for overcoming them.

Research on eliciting detailed accounts from cooperative individuals is predominantly focused on recall of isolated incidents, such as witnessing a robbery or assault. However, there are occasions where information needs to be reported about multiple repeated events, such as attendance at regular meetings of a terrorist cell or a criminal gang. Research indicates that we remember unique and repeated events differently and, as a result, reporting repeated events can be challenging.

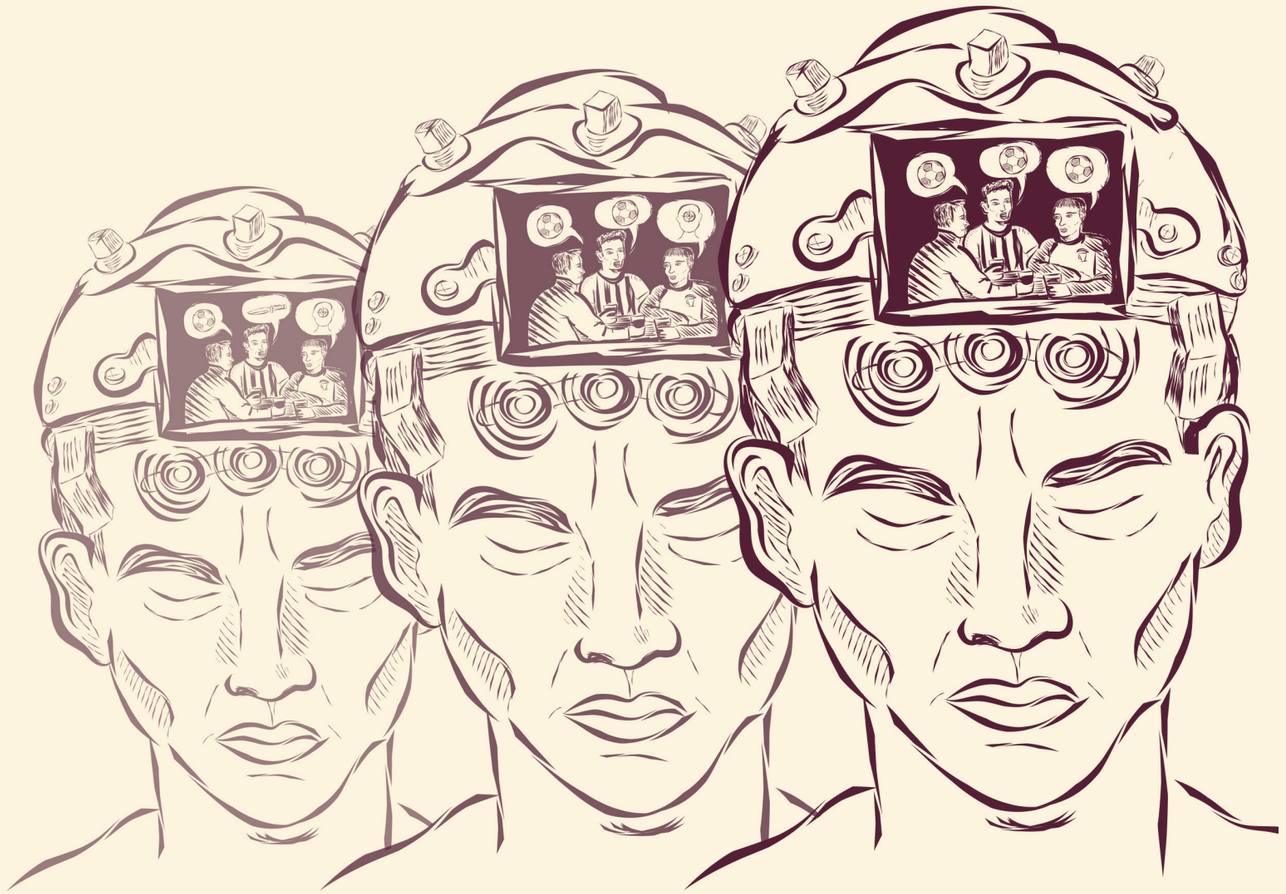
Due to the reconstructive nature of memory, when experiencing repeated events, we build a script or schema in our memory based on what usually happens.

This is an adaptive mechanism of memory that we use every day, so that we anticipate what is likely to happen next time we attend a meeting at work or have a family dinner during the holidays.

The events are not necessarily identical, but there is a general – and similar – routine. In a scenario where a terrorist group plans a series of attacks, members of the group will have to meet, discuss the target and access equipment to carry out the attack. They will have to arrange the logistics for the day of the attack and allocate different responsibilities and roles to different individuals.

Although some details will predictably vary every time, for instance the target, the equipment acquired, and potentially even the main actors, the overall routine remains relatively stable across all gatherings. As a result, the variable details tend to fade away from memory faster and become part of the overall consistent routine. Thus, exposure to repeated events can be beneficial as it strengthens the script in our memory.

However, relying on the overall script and thus reporting details that are relatively fixed across events negatively affects the reporting of details which are specific to individual occurrences. If asked about several work meetings one attended in the past month, people are likely to report the gist of what usually occurs rather than something that a specific person said or did in an individual meeting. Importantly, however, these specific variable details can still be accessed provided appropriate cues are used at retrieval.



Recent evidence suggests that, if something unexpected occurred in an event, then that particular instance becomes more memorable (a targeted effect). It may even be true that the memory for all instances is improved (a general effect) as a result. For instance, if during the successful planning of a series of terrorist attacks, a specific attempt fails because of some complication that would be a deviation from the general script, this might enhance one's memory for the unsuccessful attempt or for the whole series of attacks – to some extent because it facilitates separating the repeated events.

To date, research on effective memory-enhancing techniques to facilitate reporting of repeated events with adult witnesses or informants is very limited. However, there is a wealth of research on improving the reporting of individual instances of repeated events with child interviewees, aiming to facilitate criminal investigations of cases of abuse.

Evidence suggests that to improve recall for instance-specific details, aiding recall of individual instances is crucial. To this end, a strategy that would be consistent with what we know about memory is to initially encourage a free narrative about the events to facilitate accurate reporting and the identification of 'labels' for individual occurrences.

A related strategy is to ask about a time that was more memorable from the series of events, or given that a deviation might be more memorable, to ask whether there was a time where something different happened. Crucially, although interviewers are encouraged to ask instance-specific questions to elicit information about individual occurrences, they should use open-ended prompts as there is an increased risk of source confusion in the reporting of repeated events that can consequently increase the risk for inaccurate and suggestible reporting.

While the above can help elicit detail from repeated events, there remain significant challenges and unaddressed questions regarding the use of common information elicitation techniques in these situations.

In our laboratory, we have recently conducted two experiments where we examined the effectiveness of the Timeline Technique extended by additional mnemonics and follow-up open questions to aid recall of repeated events and elicit detailed reports with adult interviewees. The Timeline, bolstered by cues and prompts, facilitated recall for specific occurrences and improved the reporting of attributions of statements and actions by perpetrators ('who did/said what and when') compared to a free request of information.

Although further research is needed in this area, our results show that the use of flexible formats that promote interviewee-led reporting can be useful in eliciting detailed accounts of complex repeated events.

Feni Kontogianni is a Postdoctoral researcher at the University of Portsmouth, working on techniques that enhance information elicitation in security contexts.

PAMELA HANWAY

COGNITIVE LOAD AT INTERVIEW: THE INTERVIEWER'S PERSPECTIVE

Psychological research has, for many years, provided practitioners with guidance on best practice for interviewing witnesses and suspects. Advice has also been provided, for intelligence-gathering practitioners, regarding the retrieval of information. To assist interviewers, several techniques have been developed for use in a diverse range of information-gathering settings. These include the PEACE protocol, cognitive interviewing, and best practice for interviewing children and vulnerable witnesses, e.g., Achieving Best Evidence (ABE).

However, despite guidance and training interviewers often do not, or perhaps cannot, comply with the guidance. This can have serious consequences for individuals and the wider context, such as the Criminal Justice process. So why is compliance with best practice difficult and what makes investigative interviewing so demanding?

COGNITIVE LOAD FOR INTERVIEWERS

One factor is the effect of cognitive load on the performance of interviewers. 'Cognitive load' encapsulates a wide variety of terms used to describe the phenomenon of working memory use and includes cognitive workload, mental strain and the mental effort required to complete tasks. We all have a relatively limited cognitive capacity to perform simultaneous tasks and cognitive overload may result, thereby affecting performance.

Research has shown that increasing cognitive load impacts interviewees in terms of their retrieval of information. However, the effect for interviewers, in forensic settings, has not been examined. What we do know, is that in other applied settings, for example interviewing for workplace recruitment, cognitive load can have an impact upon decision making.

Cognitive load may also influence the performance of airline pilots, air traffic controllers, and medical trainees. For example, when trainee surgeons perform a cognitively demanding surgical procedure there can be a negative impact upon their performance. Cognitive load, therefore, may have serious consequences when it comes to intelligence-gathering in high-stakes situations.

In investigative interviews, there are several cognitive processes occurring simultaneously for interviewers. They are required to actively listen to their interviewees and to remember information provided. The information needs to be processed, assimilated and considered along with knowledge interviewers may already possess, or which is passed to them during the course of an interview. Interviewers have to make reasoned judgments, formulate appropriate questions and decide upon their responses. However, their limited capacity to process information could lead to cognitive overload, which may impact upon interviewers' performance, making the process of obtaining accurate and detailed accounts more difficult.

INTERVIEWING IN THE 'REAL-WORLD'

In my research, we assessed the impact of cognitive load on officers from two UK police forces, who had been trained in various interview techniques. When interviewing they expressed that it was cognitively demanding, stating for example, 'you're thinking hang on a minute, slow down, I've got to remember all this', and explained that the cognitive load they experience sometimes impacted upon their performance, "if you haven't identified the right thing in an interview it can have a massive effect". Analysis of the interviewers' experiences identified key features of interviewing that may increase cognitive load.

These triggers of cognitive load included time pressures due to operational requirements and specific aspects of the interview task, for example, withholding information from interviewees and the formulation of appropriate questions. They also identified areas of planning and preparation, or a lack thereof, as being significantly detrimental to their performance.

REDUCING THE COGNITIVE BURDEN

Cognitive load, therefore, can result from a combination of task characteristics, such as time pressure and complexity. Ensuring that sufficient time is allowed for the interviewer to conduct the interview and undertaking effective planning and preparation, particularly for complex or challenging investigations, can reduce cognitive load. As a consequence, managing the interview task in this way may enhance compliance with best practice guidance, as well as increasing the quantity and quality of information gained.

Pamela Hanway is a PhD student at the University of Portsmouth, her current research focuses on the effects of cognitive load for investigative interviewers. Pamela was formerly a detective within a UK police force and has a wealth of investigative and interviewing experience.

SARAH MARSDEN

COUNTERING VIOLENT EXTREMISM: A GUIDE TO GOOD PRACTICE

Since the early 2000s, more than fifty countries have developed initiatives to counter violent extremism (CVE). Despite this, there still remains a lack of strong evidence on which interventions are effective. With colleagues James Lewis and Kim Knott, Sarah Marsden has reviewed the literature on CVE programmes, to give examples of what good CVE practice should look like.

CVE takes many different forms, from government-led programmes such as those included in the UK's Counter-Terrorism strategy to grassroots initiatives – as outlined by our colleague Ben Lee (*CSR*, Issue 3), a wide range of actors and approaches fall under efforts to counter violent extremism.

Because the factors which lead to violent extremism are complex and wide-ranging, the content of programmes to counter it are diverse. Consequently, the scope and definition of CVE initiatives can be unwieldy. For example, the European Commission, in 2015, defined CVE as 'all actions that strengthen the resilience of individuals and communities to the appeal of radicalisers and extremism'. With such broad definitions, it can often be unclear how some programmes, categorised as 'CVE-relevant', can be seen to impact on violent extremism.

Despite this, after over a decade of CVE initiatives a useful picture has begun to emerge, and while there is a strong need for research and evaluation on the impact of CVE programmes, we can begin to point towards evidence of good practice relating to the design, delivery and assessment of some initiatives.

PROGRAMME DESIGN

There is increasing awareness of the need to carefully target CVE programmes, so they are directed at different stages of the journey into and out of extremism.

- Primary interventions have the broadest scope. These target whole sections of a community in an effort to raise awareness about extremism and try to address its 'root causes'.
- Secondary interventions engage with those considered at risk of involvement in extremism, aiming to disrupt the process of radicalisation.
- Tertiary interventions are concerned with individuals already involved in extremism and seek to support disengagement, deradicalisation, and reintegration.

PROGRAMME DELIVERY

A wide range of actors are involved in developing and delivering CVE interventions. Some programmes are highly centralised, and are run and managed by central and local government, others are instigated by civil society actors such as faith or community organisations, NGOs, or former combatants. International bodies such as the European Union are also involved in CVE work.

The extent of involvement from different actors varies; however, most interventions reflect a hybrid approach involving some form of cooperation between government and local actors.

These collaborative efforts are better able to address the complex dynamics of violent extremism, but need to ensure they don't undermine the legitimacy of community-based groups perceived to be working too closely with government.

PROGRAMME EVALUATION

The evidence base about what works in CVE is weak. Few programmes conduct systematic evaluations and many don't make their assessments public. There is also little agreement on what looks like success and how to measure outcomes.

Evaluation can be achieved through three differing approaches. A common approach is by interpreting change in risk factors which operate across a number of levels, including personal factors, such as a desire for adventure or belonging, or need for status; political influences, including a sense of grievance, or strong identification with a political or religious ideology; and group dynamics, such as family or peer involvement in extremism.

To take account of the wider context within which reintegration takes place, it can be helpful to supplement risk-oriented measures by interpreting how well someone is reintegrating. This can include economic integration, such as employment, education or training; social integration, including positive relationships with friend or family networks that do not support extremism; and political integration, such as engagement with democratic systems and increased commitment to wider social and political norms.

Another method of assessing interventions is by examining the process by which organisations develop and deliver their programmes. These can include measures which determine the programme's integrity, including whether a programme's aims relate to its methods and outcomes and the strength of the evidence that supports this theory of change; delivery agents, including the degree of legitimacy and credibility an intervention provider holds in the local community; and multi-agency working, such as the scope of relationships with relevant statutory and non-statutory organisations and the degree to which the intervention is able and willing to engage with existing multi-agency collaborations.

GOOD PRACTICE IN CVE

In a guide published by CREST, we provided case-studies that draw out examples of these methods of evaluation and where and how evaluations have worked, or not. Drawing on these studies, and our evaluation of the wider CVE literature, we have identified several important points to take account of when designing, implementing and evaluating CVE programmes. It is helpful to consider what constitutes a successful outcome of an intervention, and how this might be assessed and communicated. To help with this, it is important to determine the boundaries of what is CVE-relevant by clarifying which causes of violent extremism interventions are seeking to address and specifying the mechanism by which they are designed to work.

Interventions should also balance a structured approach with the flexibility necessary to respond to unexpected events and shifting local needs. In addition, their design should be based on empirical evidence that informs a theory of change linking aims, methods, and outcomes. Governments have an important role in designing, funding, and assessing CVE initiatives, as well as building the capacity of community-based actors. Capacity building is helped through fostering local support for interventions by engaging with a range of relevant local and national agencies and stakeholders. Working with community-based partners and families helps in understanding local context, as well as demonstrates credibility and legitimacy in ways that government programmes can find difficult.

Our guide on CVE provides a range of intervention models which reflect different aspects of good practice in their design, delivery and assessment. Whilst ongoing research and evaluation is undoubtedly a priority, there is much to learn from existing practice.

Dr Sarah Marsden is Lecturer in Radicalisation and Protest in the Department of Politics, Philosophy and Religion at Lancaster University. Her book, Reintegrating Extremists: Deradicalisation and Desistance *is available with Palgrave MacMillan. With Professor Kim Knott and James Lewis she has written a guide to good practice on CVE.*

JOEL BUSHER, DONALD HOLBROOK AND GRAHAM MACKLIN

EXPLAINING NON- OR LIMITED ESCALATION OF VIOLENCE: THE ROLE OF 'INTERNAL BRAKES'

Why do some 'extremists' or 'extremist groups' choose not to engage in violence, or engage only in particular forms of low-level violence? Why, even in deeply violent groups, are there often thresholds of violence that members rarely if ever cross?

Part of the answer is likely to lie in external constraints, such as the counter-measures put in place by state and non-state actors to inhibit the activities of such groups. Yet the fact that few if any groups carry out as much violence as they are capable of, indicates that in most cases external constraints comprise only part of the answer. Detailed empirical accounts indicate that pressures within these groups also inhibit the adoption or diffusion of greater violence. In other words, the limits on violence are to some extent self-imposed. To date, however, there has been little systematic analysis of these 'internal brakes' on violent escalation.

In response to this gap in understanding, we set out to develop a typology to describe and categorise the internal brakes on violent escalation within extremist groups – including both more and less formalised groups. We drew three broad conclusions.

1. A single typology of the internal breaks on violent escalation can have applicability across groups characterised by different ideologies and levels of violence.

We developed and tested the typology using three primary case studies that differed significantly in terms of ideology and levels of violence: the transnational and British jihadi scene from 2005 to 2016; the British extreme right during the 1990s, and the animal liberation movement in the UK from the mid-1970s until the early 2000s. This made it possible to test if the typology could be applied to different actors.

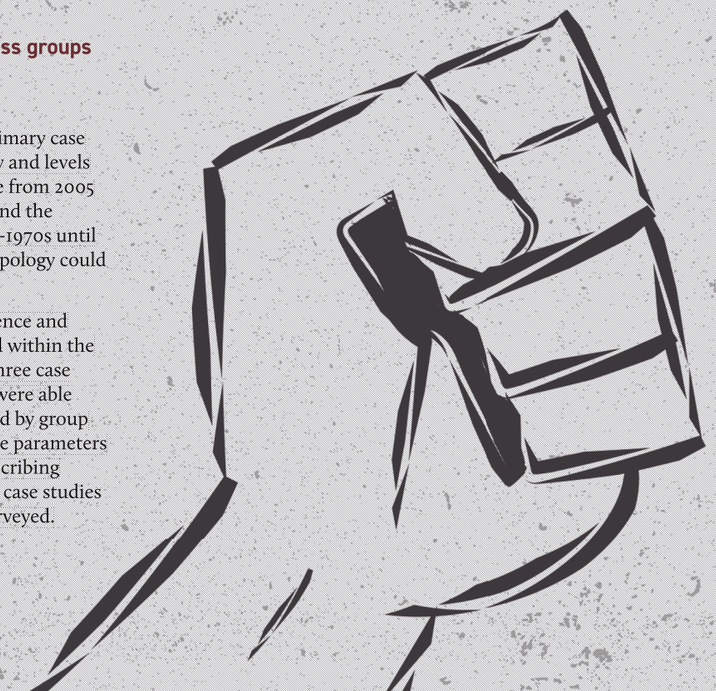
As expected, we found that the distribution, prominence and effectiveness of brakes varied considerably across and within the three primary case studies. Nonetheless, across the three case studies and across the wider literature surveyed, we were able to (a) identify broadly similar practices being deployed by group members as they sought to establish and maintain the parameters of their violence, and (b) develop a vocabulary for describing these practices that could be applied across the three case studies and to other examples drawn from the literatures surveyed.

2. The internal breaks on violence escalation appear to operate as a series of underlying logics.

While the analysis revealed a wide array of practices through which group members seek to establish and maintain parameters on their own group's violence, we found that these operate on five basic underlying logics. For each of these logics, we identified a higher order brake and a series of sub-brakes, as summarised in the table opposite.

Organising the typology in this way has two main advantages. By reducing the typology down to five high-level categories it provides a manageable system of categorisation.

More importantly, it also helps to reveal how different brakes work and, by extension, can provide insight about how different brakes can either reinforce or contradict one another.



3. While the typology opens up some potentially productive avenues of research and analysis, it should be handled with care.

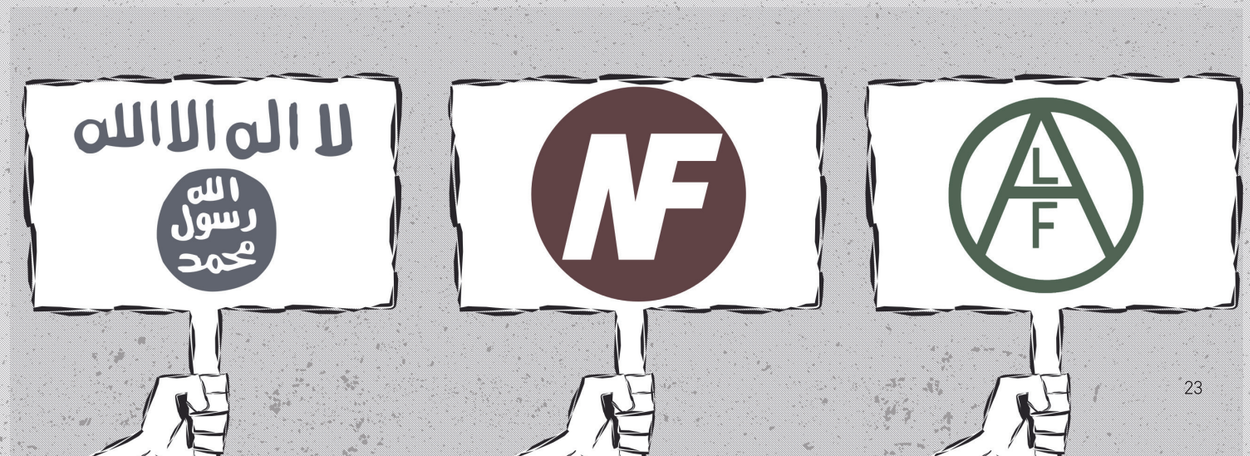
A number of issues require attention if this typology is to be used to support the assessment of the threats from, and opportunities to inhibit, the risk of escalation towards violence. Foremost among these is the fact that it cannot be use as a straightforward 'checklist'. The presence of internal brakes within any given case might be telling us one of a number of different things: it might indicate a limited risk of violent escalation due to extensive intra-movement opposition to such escalation; but it might also indicate that there are increasingly active attempts within the movement to escalate violence (hence increased 'braking'); or that there are growing intra-movement tensions. We believe nonetheless that, when used with due caution, the vocabulary that the typology provides can enhance in a number of ways the ability of researchers and analysts to investigate and understand hitherto under-researched processes of non- or limited escalation.

For researchers, the typology sets up a number of questions that are ripe for enquiry. For example: Under what conditions are certain brakes, or configurations of brakes, more likely to be effective? How are the patterns and functioning of internal brakes affected by wider conflict dynamics and vice versa? And how do the internal brakes on violent escalation operate at different points within waves or cycles of conflict?

For practitioners working in areas of risk assessment, it can provide a tool with which to identify indicators of the propensity towards and away from potential violence by groups or sub-groups. Meanwhile, for practitioners undertaking interventions with extremist groups, this typology can be used to inform assessments about how externally applied counter-measures might interact with, and sometimes undermine, internal brakes.

Dr Joel Busher is a Senior Research Fellow at the Centre for Trust, Peace and Social Relations, Coventry University, UK, Dr Donald Holbrook is an Honorary Senior Research Associate at University College London, UK and Dr Graham Macklin is a Postdoctoral Fellow at the Center for Research on Extremism (C-REX) at University of Oslo, Norway. *This research was commissioned by the* Centre for Research and Evidence on Security Threats (CREST)*, more information on the project can be found at https://crestresearch.ac.uk/projects/internal-brakes-violent-escalation.*

LOGIC	BRAKE
STRATEGIC LOGIC: Addressing questions of 'what works?'	1 Identification of non- or less violent strategies of action as being as or more effective than more violent alternatives
MORAL LOGIC: Addressing questions about whether it is 'right' or 'appropriate' to use particular forms of violence against particular targets	2 Construction of moral norms and evaluations that inhibit certain forms of violence and the emotional impulses towards violence
LOGIC OF EGO MAINTENANCE: Relating to group members' construction and maintenance of their self-image	3 Self-identification as a group that is either nonviolent or uses only limited forms of violence
LOGIC OF OUTGROUP DEFINITION: Relating to how group members conceive of their opponents and their relationship to them	4 Boundary softening in relation to putative out-groups e.g., opponents, opponents' perceived supporters, the general public or state actors
ORGANISATIONAL LOGIC: Relating to the way that organisational developments condition decision making e.g., through forms of organisational path dependency	5 Organisational developments that either (a) alter the moral and strategic equations in favour of non- or limited violence, (b) institutionalise less violent collective identities and/or processes of boundary softening, and/or (c) reduce the likelihood of unplanned violence



DUNCAN HODGES

A-Z OF DATA

ARTIFICIAL NEURAL NETWORKS
A framework for building machine learning algorithms that is inspired by the brain.

BIG DATA
Data that cannot be efficiently analysed using conventional means, typically because of its volume, veracity, velocity and/or variety.

CONVOLUTIONAL NEURAL NETWORKS
A particular type of Deep Learning that is adept at dealing with images, speech and text.

DEEP LEARNING
An approach using very complex, multi-layered Artificial Neural Networks that requires large amounts of training data but can perform very complex tasks such as image labelling, like identifying cars in a photograph.

EXPLORATORY DATA ANALYSIS
The preliminary investigations of a data set in order to better understand its characteristics.

FEATURE SELECTION
Or feature engineering, the process of selecting inputs to an algorithm and how these inputs should be represented. For example, if we are trying to create an algorithm to predict how many free seats there are on a train journey – what is the best set of information about the journey, is it start time, end time, date, starting station, destination station, what colour the train is, or the weather?

GENETIC ALGORITHM
Process that mimics natural selection, where a solution evolves through the mixing, or 'breeding', and 'mutation' of a set of potential solutions. Most often used in robotic problems or problems where there are a large number of good solutions, and we are trying to find the best from these.

HASHING
A mathematical process that takes data of any size and maps it to data of a fixed size. The process is generally difficult to reverse and is most commonly used in the storage of sensitive data such as passwords or in index structures. Normally seen in action turning passwords like 'Hunter2' into '*****'.

INDEX
A structure that allows the efficient location of a piece of information in a data store.

JUPYTER NOTEBOOK
A document that contains live code, analysis and descriptive text, allows sharing and collaboration around a data analysis task.

KOLMOGOROV-SMIRNOV TEST
A mathematical approach to analysing two datasets to determine if they have equal distributions. Helps with understanding whether two groups in an experiment show different responses to a stimulus.

LOGISTIC REGRESSION
A model that typically predicts a binary outcome (e.g., true / false) from one or more continuous inputs, such as predicting whether someone will repay a loan based on their income.

MACHINE LEARNING
The field of study dealing with algorithms and models that improve their performance as they are provided with more data. This improvement continues until overfitting occurs and maximum performance has been reached.

NATURAL LANGUAGE PROCESSING
A field of study which attempts to train machines to understand and analyse human languages, contributing to applications such as automated customer services assistants on websites.

OVERFITTING
The scourge of modern data science, generally occurs when a system has been 'over trained' on a training data set and cannot generalise to data to which it has not been previously exposed.

PRIVACY
The expectation that personally identifiable information or other sensitive data will be treated securely and sensitively. Getting value from data whilst respecting the privacy of data subjects is the cornerstone of modern data protection laws.

QUALITATIVE DATA
Data that are non-numerical in form.

'R'
A leading free software environment widely used for data analysis tasks.

SUPERVISED LEARNING
The process of learning from a set of labelled data. Typically used in classification techniques where we want to sort inputs into a number of different classes (e.g., email spam / not-spam).

TRUST
Within data science the perception of the credibility of a piece of data, a data source, a data processing system or a prediction.

UNSUPERVISED LEARNING
The process of learning where no previous data is used. Typically used in clustering techniques where we wish to group input data to a number of groups that exhibit similar characteristics, such as grouping movies into genres on a streaming platform.

VISUALISATION
The process for communicating complex information typically through imagery.

WORD EMBEDDINGS
A set of statistical Natural Language Processing techniques where words are allocated a vector of numbers. This vector of numbers effectively encodes the 'meaning' of the word. Machines can then use these vectors to better 'understand' a corpus of text.

X-AXIS
The horizontal axis on a graph, also called the abscissa – a term used at least since the 13th century, by Leonardo of Pisa.

YOTTABYTE
One septillion bytes or 1 trillion Terabytes – about 200,000 trillion photos of Kim Kardashian (see *CSR*, Issue 5)!

ZIPF'S LAW
A feature of all natural languages where the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc.

Dr Duncan Hodges is a Senior Lecturer in Cyberspace Operations at Cranfield University and is based at the Defence Academy of the United Kingdom. He holds an ESRC National Centre for Research Methods fellowship investigating Digital Identity and is a visitor at the Alan Turing Institute, the UK national institute for data science and artificial intelligence. His research focuses on identity in online and offline spaces, operations in cyberspace and how they can be supported by the innovative and ethical use of data.

WANT TO READ MORE ABOUT SOME OF THE RESEARCH THAT OUR CONTRIBUTORS MENTIONED IN THEIR ARTICLES? TAKE A LOOK BELOW. WE'VE FLAGGED UP THOSE THAT ARE OPEN ACCESS AND GIVEN LINKS TO ONLINE VERSIONS WHERE THEY ARE AVAILABLE

RYAN L. BOYD AND PAUL KAPOOR – PSYCHOLOGICAL PROFILING AND EVENT FORECASTING USING COMPUTATIONAL LANGUAGE ANALYSIS (p6)

Michael L Birnbaum, Sindhu Ernala, Asra Rizvi, Munmun De Choudhury, John Kane. 2017. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *J Med Internet Res, 19(8): e289*. Available at: https://doi.org/10.2196/jmir.7956

Ryan L. Boyd. 2017. Psychological text analysis in the digital humanities. In S. Hai-Jew (Ed.), *Data analytics in the digital humanities* (pp. 161–189). New York: Springer International Publishing.

Ryan L. Boyd, James W. Pennebaker. 2015. Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science, 26*(5): 570–582. Available at: http://bit.ly/2OuJtmA

Ryan L. Boyd, Steven R. Wilson, James W. Pennebaker, Michal Kosinski, David J. Stillwell, Rada Mihalcea. 2015. Values in words: Using language to evaluate and understand personal values. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media* (pp. 31–40). Available at: http://bit.ly/2JLNVi7

Shuki Cohen, Arie Kruglanski, Michele Gelfand, David Webber, Rohan Gunaratna. 2018. Al-Qaeda's propaganda decoded: A psycholinguistic system for detecting variations in terrorism ideology. *Terrorism and Political Violence, 30*(1): 142–171. Available at: http://bit.ly/2CHMjQu

Kimberly Glasgow, Clayton Fink, Jordan Boyd-Graber. 2014. "Our grief is unspeakable": Automatically measuring the community impact of a tragedy. In *Eighth International AAAI Conference on Weblogs and Social Media*. Available at: http://bit.ly/2FLc6YX

Vlad Niculae, Srijan Kumar, Jordan Boyd-Graber, Cristian Danescu-Niculescu-Mizil. 2015. Linguistic harbingers of betrayal: A case study on an online strategy game. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1650–1659). Beijing, China: Association for Computational Linguistics. Available at: http://www.aclweb.org/anthology/P15-1159

James Pennebaker. 2013. *The secret life of pronouns: What our words say about us* (Reprint edition). New York: Bloomsbury Press.

Matteo Vergani, Ana-Maria Bliuc. 2018. The language of new terrorism: Differences in psychological dimensions of communication in Dabiq and Inspire. *Journal of Language and Social Psychology*. Available at: https://psyarxiv.com/xg4f3/

Laura Wendlandt, Rada Mihalcea, Ryan L. Boyd, James Pennebaker. 2017. Multimodal analysis and prediction of latent user dimensions. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I* (Vol. 10539 LNCS, pp. 323–340). Cham: Springer International Publishing. Available at: http://bit.ly/2TYR9U2

PIP THORNTON – WORDS AS DATA (PAGE 8)

Pip Thornton. 2017. Geographies of (con) text: language and structure in a digital age. *Computational Culture*, Issue 6. Available at: http://bit.ly/2WtsNi6

Pip Thornton. 2018. A critique of linguistic capitalism: provocation/intervention. *GeoHumanities*, 4 (2): 417-437. Available at: http://bit.ly/2WwluX4

Pip Thornton. 2019. *Language in the age of algorithmic reproduction: A critique of linguistic capitalism*. PhD. Royal Holloway, University of London. Available at: http://bit.ly/2UfHckn

Pip Thornton. 2019. Linguistic Geographies – Language in the Age of Algorithmic Reproduction. Website at: https://linguisticgeographies.com.

JO HINDS – BEHAVIOUR PREDICTION (p10)

Adam Bermingham, Alan Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*: 2-10. Available at: https://core.ac.uk/download/pdf/11310530.pdf

Philip Howard, Muzammil Hussain. 2013. *Democracy's fourth wave? Digital media and the Arab Spring*. Oxford: Oxford University Press.

Márton Mestyán, Taha Yasseri, János Kertész. 2013. Early prediction of movie box office success based on Wikipedia activity big data. *PloS one, 8*(8): e71226. Available at: http://bit.ly/2ClxVrg

Amy Nodrum. 2016. Popular Internet of Things forecast 50 billion devices by 2020 is outdated. *IEEE Spectrum*, 18. Available at: http://bit.ly/2Fysfkj

DEBI ASHENDEN – ALGORITHMIC DECISION MAKING (p12)

D. Elliott Bell, Leonard LaPadula. 1973. Secure computer systems: Mathematical foundations (No. MTR-2547-VOL-1). MITRE Corp. Available at: http://bit.ly/2CHV9O5

Lizzie Coles-Kemp, Alf Zugenmaier, Makayla Lewis. 2014. Watching you watching me: The art of playing the panopticon. *Digital Enlightenment Yearbook 2014: Social Networks and Social Machines, Surveillance and Empowerment*, 147. Available at: http://bit.ly/2Fz0sA1f

Lizzie Coles-Kemp, René Hansen. 2017. Walking the line: The everyday security ties that bind. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*. 464-480. Available at: http://bit.ly/2WrYng7

Lizzie Coles-Kemp, Debi Ashenden, Kieron O'Hara. 2018. Why Should I? Cybersecurity, the security of the State and the insecurity of the citizen. *Politics and Governance, 6*(2): 41-48. Available at: http://bit.ly/2CFnUew

Michael Kearns, 2017. Fair algorithms for machine learning. In *Proceedings of the 2017 ACM Conference on Economics and Computation*. Available at: https://dl.acm.org/citation.cfm?id=3084096

Alvin Roth. 2007. Repugnance as a constraint on markets. *Journal of Economic Perspectives*. 21(3): 37–58. Available at: https://www.nber.org/papers/w12702

Alvin Roth. 2008. What have we learned from market design? *The Economic Journal*, 118: 285–310. Available at: https://www.nber.org/papers/w13530

Ravi Sandhu, Edward Coyne, Hal Feinstein,Charles Youman. 1996. Role-based access control models. *Computer*, 29(2): 38-47. Available at: http://bit.ly/2HM47xS

PAUL TAYLOR – FROM DATA TO DATUM (p14)

Galit Nahari, Tzachi Achkenazi, Ron Fisher, et al. 2019. Language of lies: Urgent issues and prospects in research. *Legal and Criminological Psychology*, 24: 1-23. doi:10.1111/lcrp.12148. Includes comment: Paul Taylor, Abbie Marono, Lara Warmelink. 2019. The ecological challenge: Ensuring our aggregate results are individually relevant. *Legal and Criminological Psychology*, 24: 4-8. Available at: http://bit.ly/2uyZkr8

Paul Taylor, et al. Coral Dando, Tom Ormerod, Linden Ball, Marisa Jenkins, Alexandra Sandham, Tarek Menacere. 2013. Detecting insider threats to organizations through language change. *Law and Human Behavior*, 37: 267-275. Available at: http://bit.ly/2JL3HcV

Information from the CREST project, led by Professor Ashraf Labib on 'Taking Decisions about Information Value' is available from: http://bit.ly/2JW1tI6

PAMELA HANWAY – COGNITIVE LOAD AT INTERVIEW (p18)

Roger Dias, M.C. Ngo-Howard, M. T. Boskovski, Marco Zenati, S. J. Yule. 2018. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *British Journal of Surgery*. Available at: http://doi.org/10.1002/bjs.10795

Rachel Frieder, Chad Van Iddekinge, Patrick Raymark. 2016. How quickly do interviewers reach decisions? An examination of interviewers' decision-making time across applicants. *Journal of Occupational and Organizational Psychology*, 89(2): 223–248. Available at: http://bit.ly/2Wvry20

Edith Galy, Magali Cariou, Claudine Mélan. 2012. What is the relationship between mental workload factors and cognitive load types?

International Journal of Psychophysiology, 83(3), 269-275. Available at: http://bit.ly/2HYz8Op

Pamela Hanway, Lucy Akehurst. 2018. Voices from the front line: police officers' perceptions of real-world interviewing with vulnerable witnesses. *Investigative Interviewing: Research and Practice*. Available at: https://www.iiirg.org/journal/volume-9-issue-1/

Gavin Oxburgh, James Ost, Paul Morris, Julie Cherryman. 2015. Police officers' perceptions of interviews in cases of sexual offences and murder involving children and adult victims. *Police Practice and Research*, 16(1): 36-50. Available at: http://bit.ly/2Ottaqh

SARAH MARSDEN – COUNTERING VIOLENT EXTREMISM (p20)

Toke Agerschou. 2014. Preventing radicalization and discrimination in Aarhus. *Journal of Deradicalization*, 1: 5-22. Available at: http://bit.ly/2v2FSia

Daniel Aldrich. 2012. Radio as the voice of God: Peace and tolerance radio programming's impact on norms. *Perspectives on Terrorism* 6(6): 34-60. Available at: http://bit.ly/2CDmqB9

James Khalil, Martine Zeuthen. 2014. *Qualitative study on countering violent extremism (CVE) programming under the Kenya Transition Initiative*. USAID. Available at: http://bit.ly/2uCCP4p

Daniel Koehler. 2013. Family counseling as prevention and intervention tool against the German 'Hayat' program. *Journal Exit Germany. Journal of Deradicalization and Democratic Culture*, *3*: 182-204. Available at: http://bit.ly/2Oyr1cB

Jon Kurtz, Rebecca Wolfe, Beza Tesfaye. 2016. *Does youth employment build stability? Evidence from an impact evaluation of vocational training in Afghanistan*. In Sara Zeiger (Ed.). Expanding research on countering violent extremism. Hedayah. Available at: http://bit.ly/2l298lz

Jose Liht, Sara Savage. 2013. Preventing violent extremism through value complexity: Being Muslim being British. *Journal of Strategic Security*, 6(4): 44-66. Available at: https://scholarcommons.usf.edu/jss/vol6/iss4/3/

Lilla Schumicky-Logan. 2017. Addressing violent extremism with a different approach: The empirical case of at-risk and vulnerable youth in Somalia. *Journal of Peacebuilding & Development*, 12(2): 66-79. Available at: http://bit.ly/2YttRVb

Michael Williams, John Horgan, William Evans. 2016. *Evaluation of a multi-faceted, U.S. community-based, Muslim-led CVE program*. Washington, DC: National Institute of Justice. Available at: http://bit.ly/2HLkzyz

JOEL BUSHER, DONALD HOLBROOK, GRAHAM MACKLIN – EXPLAINING NON OR LIMITED ESCALATION OF VIOLENCE (p22)

Tore Bjørgo, Ingvild Gjelsvik. 2017. Right-wing extremists and anti-Islam activists in Norway: Constraints against violence. *C-REX Working Paper Series*, No.3. Available at: http://bit.ly/2HYzMM6

Joel Busher, Donald Holbrook, and Graham Macklin. 2019. The internal brakes on violent escalation: A typology. *Behavioral Sciences of Terrorism and Political Aggression* 11 (1): 3–25. Available at: http://bit.ly/2OyoZX5

R. Kim Cragin. 2014. Resisting violent extremism: A conceptual model for non-radicalization. *Terrorism and Political Violence*, 26(2): 337-353. Available at: http://bit.ly/2JNgw6z

Martha Crenshaw. 1996. Why violence is rejected or renounced: A case study of oppositional terrorism. In T. Gregor (Ed.), *A natural history of peace* (pp. 249-272). Nashville, TN: Vanderbilt University Press.

Maiah Jaskoski, Michael Wilson, Berny Lazareno. 2017. Approving of but not choosing violence: Paths of nonviolent radicals. *Terrorism and Political Violence*. Available at: http://bit.ly/2V29xl4

Emmanuel Karagiannis, Clark McCauley. 2006. Hizb ut-Tahrir al-Islami: Evaluating the threat posed by a radical Islamic group that remains nonviolent. *Terrorism and Political Violence*, 18(2): 315-334. Available at: http://bit.ly/2HKYxvO

Iona Emy Matesan. 2018. Organizational dynamics, public condemnation and the impetus to disengage from violence. *Terrorism and Political Violence*. Available at: http://bit.ly/2l3xNG2

Pete Simi, Steven Windisch. 2018. Why radicalization fails: Barriers to mass casualty terrorism. *Terrorism and Political Violence*. Available at: http://bit.ly/2Wr6V6Y

Maria Stephan, Erica Chenoweth. 2008. Why civil resistance works: The strategic logic of nonviolent conflict. *International Security*, 33(1): 7–44. Available at: http://bit.ly/2CDmUqX

CREST CENTRE FOR RESEARCH AND EVIDENCE ON SECURITY THREATS

CREST Security Review provides a gateway to the very best knowledge and expertise. Its articles translate academic jargon to 'so what' answers and illustrate how behavioural and social science can be used effectively in everyday scenarios.

THE CENTRE FOR RESEARCH AND EVIDENCE ON SECURITY THREATS

CSR is produced by the Centre for Research and Evidence on Security Threats (CREST). CREST is funded by the UK's security and intelligence agencies to identify and produce social science that enhances their understanding of security threats and capacity to counter them. CREST also receives funding from its core partners (the universities of Bath, Lancaster and Portsmouth). Its funding is administered by the Economic and Social Research Council (ESRC Award ES/N009614/1), one of seven UK Research Councils, which direct taxpayers' money towards academic research and training. The ESRC ensures the academic independence and rigour of CREST's work.

CREST has established a growing international network of over 100 researchers, commissioned research in priority areas, and tackled some of the field's most pressing questions.

> **There really is some impressive work going on. Yet, all that effort is irrelevant if practitioners, policy-makers, and other stakeholders do not get to hear about it. *CREST Security Review* is one way we will keep stakeholders informed not only on what CREST is doing, but also on the best research from around the world.**
> **Professor Paul Taylor, CREST Director.**

For more information on CREST and its work visit **www.crestresearch.ac.uk** and find us on Twitter, Facebook and LinkedIn.

Lancaster University · University of Portsmouth · University of Bath · ESRC Economic & Social Research Council

A NEW HOME AND APP FOR CREST SECURITY REVIEW

CREST is constantly looking to improve and innovate in the way it translates and communicates leading research from the behavioural and social sciences on security threats. With that in mind, we've been working on making this magazine, *CREST Security Review*, more accessible to its readers.

Alongside the mobile application, we've launched a new website to give *CSR* its own home, making it easier to access, read and share articles. Favourites and bookmarks are synced with the application. Visit:

www.crestsecurityreview.com

The free app, available on both iOS and Android platforms, provides a number of features:

Intuitive navigation
Making it easy to browse issues and articles of interest.

Offline access
Read articles or whole issues by bookmarking them to read later.

Search and discover
Navigate easily to the topics that interest you most as well as be alerted when new items are added.

Favourite articles
Our synced bookmark feature allows you to conveniently store all your favourite *CSR* articles in one place, for you to access later via any of your devices.

Responsive design
You can read *CSR* on your mobile, iPad, Android tablet or desktop, meaning you no longer must download the pdf (although that option is still available).

New issue alerts
Opt in for notifications and you'll be the first to know when the latest issue arrives.

Upcoming features
As the app matures, we'll be adding bonus content for subscribers.

GET IT ON Google Play

Download on the App Store

DOWNLOAD THE FREE CSR APP

www.crestsecurityreview.com/appdownload