Title
Mundane Is the New Radical

Permalink
https://escholarship.org/uc/item/6g81c2wn

Journal
IEEE Technology and Society Magazine, 37(2)

ISSN
0278-0097

Authors
Shirley, Rebekah
Kammen, Daniel

Publication Date
2018-06-01

DOI
10.1109/mts.2018.2826076

Peer reviewed

A PUBLICATION OF THE IEEE SOCIETY ON SOCIAL IMPLICATIONS OF TECHNOLOGY

IEEE Technology and Society Magazine

VOL. 37, NO. 2, JUNE 2018

Special Section: Social Media in the Middle East

KEEPING THE LIGHTS ON

IEEE

2018 IEEE International Symposium on Technology and Society (ISTAS)

"Technology, Ethics, and Policy"

November 13 and 14, 2018
Washington, DC U.S.A.

The IEEE Society for Social Implications of Technology (SSIT) invites you to attend and contribute at our flagship annual event, the 2018 IEEE International Symposium on Technology and Society (ISTAS). ISTAS is a multi-disciplinary and interdisciplinary forum for engineers, social scientists and technologists, policy makers, scientists and entrepreneurs, and philosophers, researchers and polymaths to collaborate, exchange experiences, and discuss the social implications of technology. IEEE ISTAS 2018 is being hosted by the School of Engineering and Applied Science at George Washington University, 800 22nd Street, NW, Washington, D.C., USA.

Plenary Panels:

- A Partnership Approach to Community led Sustainable Development
- Creating Ethically Informed Standards
- Current Challenges in Technology Policy
- The Future of Ethical Education

Specialist Panels (parallel participatory sessions):

- Leveraging Geodata to Inform Public Policy
- Creating Ethically Informed Standards
- Children - Digital Rights and Risks

Parallel Panel and Workshop proposals Deadline: June 16, 2018

Call for Papers – *Deadline extended*: June 30, 2018

We welcome proposals for papers, parallel panel and workshop sessions focused on the relationship between technology, policy and social issues ranging from the economic and ethical to the cultural and environmental. Priority will be given to submissions addressing SSIT's Five Pillars and Intersections between the social implications of technology and:

- Personal Privacy vs. National Security
- Net-Neutrality
- Big Data based Decision Making
- Human Genome Editing (e.g., CRISPR)
- Ethics, neurotechnology/big brain
- Internet of Things (IoT)
- BlockChain Everything – what does it mean?
- Open Data and Open Government

Papers accepted for the Conference Proceedings will be submitted for publication in *IEEE Xplore*, with some papers selected for revision and publication in a special issue of *IEEE Technology and Society Magazine,* and potentially other journals.

http://sites.ieee.org/istas-2018/

Digital Object Identifier 10.1109/MTS.2018.2839778

Paul M. Cunningham

Operationalizing SSIT's 5 Pillars

Pillar 4: Societal Impact of Technology

This month I will briefly discuss SSIT Pillar 4, which is dedicated to Societal Impact of Technology. Pillar 4 focuses on highlighting and supporting the development of technologies that incorporate the principles of safety, security, and privacy by design.

This is an ambitious objective which requires consideration of both major themes as well as a number of major technological areas. Major themes include privacy, security, safety, ethical, legal, and political implications of current and emerging technologies and applying lessons learned from the past based on the often unintended and unanticipated implications of technological innovation and adoption.

Major technological areas that the SSIT community must address include:

- Data-driven technologies (e.g., big data analysis, Internet of Things (IoT), personalization, social networks);
- Cloud-based technologies (e.g., cloud-based computing, cloud-based transactions);
- Smart and autonomous systems (e.g., smart homes and devices, semi-autonomous and autonomous vehicles, surgical and medical robots, non-human intelligence);
- Human-centered technologies (e.g., HCI, ergonomics);
- Human enhancing technologies (e.g., augmented and virtual reality, genetics and neurology, cyber-physical systems).

It is clear that addressing this breadth and depth of technological complexity requires a human-values-based impact assessment. This can only be achieved through cross-disciplinary and interdisciplinary collaboration between engineers, technologists and scientists, medical practitioners and legal scholars, philosophers and ethicists. In the light of this requirement, Pillar 4 will prioritize multi-stakeholder collaboration across IEEE societies, as well as engagement with non-IEEE stakeholders across the public, private, education and research, and societal sectors. This is essential to develop ethical based technological, standards, and policy-related approaches to mitigate and ameliorate some of the identified safety, security, and privacy issues identified.

We hope you will be encouraged to prepare paper, panel and workshop submissions to the Call for Papers for IEEE ISTAS 2018 (Washington, DC, November 13–14), which will be addressing many of these issues.

Call for Volunteers

I invite you to help SSIT continue to make a difference in an arena of enormous complexity. Other volunteer opportunities include:

- Serving your local community through an existing or new SSIT Chapter.
- Contributing to the work of SSIT's committees (including our Standards committee).
- Volunteering to host SSIT Distinguished Lecturers.
- Submitting articles or review submissions to *IEEE Technology and Society Magazine.*
- Reviewing submissions to IEEE ISTAS, Norbert Weiner, IEEE Ethics, IST-Africa Week, and other SSIT supported conferences.
- Supporting activities of the IEEE SSIT IST-Africa SIGHT in IST-Africa Partner Countries.
- Representing SSIT on IEEE committees (TAB, BoD, Standards, Future Directions Initiative).
- Serving on the SSIT Board of Governors.

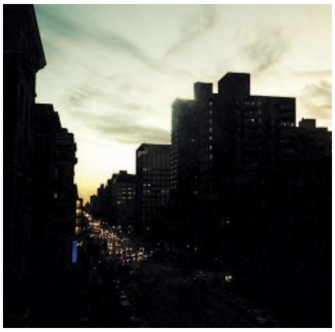
If any of these opportunities are of potential interest or if you would like to recommend someone, please contact me (Subject: Volunteer for IEEE SSIT – <name>) and I will direct you to the responsible team. If you have not received a response to

(continued on page 17)

Digital Object Identifier 10.1109/MTS.2018.2837538
Date of publication: 31 May 2018

IEEE
Technology and Society Magazine

Volume 37, Number 2, June 2018



ON THE COVER:
KEEPING THE LIGHTS ON, BY HILARY BROWN, P. 62. HIGH RELIABILITY ORGANIZATIONS CAN BE DESIGNED TO COMPENSATE FOR HUMAN FALLIBILITY AND TECHNOLOGICAL FAILURES. IMAGE: NIKKIJOHNSONPHOTO/ISTOCK.

Features

SPECIAL SECTION SOCIAL MEDIA IN THE MIDDLE EAST

* Refereed articles.

Digital Object Identifier 10.1109/MTS.2018.2826085
Date of publication: 31 May 2018

Departments

Call for Papers

President's Message

Editorial

Book Reviews

Leading Edge

Opinion

Industry View

Guest Editorial

Last Word

Special Section Articles
*Refereed articles.

IEEE SSIT SOCIETY ON SOCIAL IMPLICATIONS OF TECHNOLOGY

ieeessit.org

IEEE Technology and Society Magazine

IEEE Technology and Society Magazine, published by the IEEE Society on Social Implications of Technology, serves to facilitate understanding of the complex interaction between technology, science, and society; its impact on individuals and society in general; professional and social responsibility in the practice of engineering, science, and technology; and open discussion on the resulting issues.

A PUBLICATION OF THE IEEE SOCIETY ON SOCIAL IMPLICATIONS OF TECHNOLOGY

IEEE Technology and Society Magazine (ISSN 0278-0097) (ITSMDC) is published quarterly by The Institute of Electrical and Electronics Engineers, Inc. IEEE Headquarters: 3 Park Avenue, 17th Floor, New York, NY 10016-5997. IEEE Service Center: 445 Hoes Lane, Piscataway, NJ 08855-1331 U.S.A. $33.00 per year (digital copy included in Society fee) for each member of the IEEE Society on Social Implications of Technology. Print Subscriptions: Additional $120.00 per year for each member of the IEEE Society on Social Implications of Technology (SIT). Non-member subscription prices available on request. Single copy prices: members $30.00; nonmembers $109.00 Copyright and reprint permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. Copyright law for private use of patrons: (1) those post-1977 article that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 U.S.A.; (2) pre-1978 articles without fee. For other copying, reprint, or republication permission, write to the IEEE Copyright and Permissions Dept., IEEE Service Center. All rights reserved. Copyright ©2018 by the Institute of Electrical and Electronics Engineers, Inc. Printed in U.S.A. Periodicals postage paid at New York, NY and at additional mailing offices. **Postmaster:** Send address changes to IEEE Technology and Society Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854-4150 U.S.A. **PRINTED IN U.S.A.**

Digital Object Identifier 10.1109/MTS.2018.2826086

Jeremy Pitt

Publish or Impoverish

Academic Publishing and the Platform Economy

It can be glibly asserted that technology makes accomplishing various activities easier. But it is not always obvious for whom it makes it easier to accomplish what. For example, the Internet has had a profound impact on academic publishing, and the transition from printed paper to digital format has ostensibly made it "easier" for academics to put their work in the public domain and, if they can actually get attention in a social-media sound-bite distracted world, reach a wider audience than ever before. However, if this transition coincides — by luck or judgement — with other societal changes, then it can also make it easier for some enterprises to deploy business models that enable them to accomplish their objectives. In an ideal world, this would create a "win-win-win" scenario: a win for the academics, a win for the enterprises, and a win for society.

Thinking first of some ongoing societal changes, it is widely recognized that information and communication technology, the knowledge economy, the digital economy, etc., are profoundly important economic drivers, and that a well-educated population, as well as being a benefit in and of itself, is a prerequisite for nation states to compete in a supranational market for electronic goods and services. For such reasons, then, a country such as the United Kingdom (U.K.) sets itself a target for 50% of its 18-year olds to go to University to study for a higher degree.

Leaving aside the thorny issue of who is actually "paying" to achieve this target, which involves a considerable expansion of the sector,[1] one consequence of more students[2] is that more academics are required to teach undergraduates[3] and to "train" postgraduates. That could be seen as a beneficial outcome: after all, this is a sign surely of a well-educated population. On the other hand, it also means more academics seeking funding for their research, more academics and their students seeking publication of their research — and more proto-academics pursuing careers. Therefore the expansion has had (arguably) some less beneficial outcomes: for example, a subtle change in the nature of a Ph.D. that makes it more adversarial between supervisor and student (rather than a co-production of supervisor and student against a research question), and a diminution of the difficult transition from absorber of knowledge to creator of new knowledge,[4] the cornerstone of any Ph.D. judgement. Most unfortunately, the academics themselves have been victims of their own success, and have produced new academics in greater numbers than are needed to service this increase in demand, and to supply their own replacement. This has created an excessive pool of well-qualified and cheap labor, employable on short-term temporary or even "zero hours" contracts.

Another consequence of this expansion has been the corresponding enlargement (and indeed self-empowerment) of management and administration. In particular, there is an increased use of metrics for measuring academic contribution, and being used in appointment and promotion panels.[5] Such metrics include the

[1] Hint: the answer includes the students and the academics themselves; but not some of the primary beneficiaries of a well-educated workforce for whom tax avoidance on a, literally, industrial scale is routine.

[2] Together with legislation that practically obliges academic institutions to compete with each other to attract students, another consequence is an inexorable rise in the proportion of university budgets being spent on marketing and administration in relation to the actual teaching budget.

[3] At least until the technology of the massive open online course (MOOC) renders all but one of the teachers redundant.

[4] The pressure on completion rates shifts the burden of risk from student to supervisor, and provides less experience of "training for failure" — not every experiment will prove its null hypothesis, but that's not what one might believe if one only read Ph.D. theses.

[5] This has been referred to as the "McKinseyisation of academia" — i.e., everything can be measured, and if it can be measured then it can be managed. See also [9].

Digital Object Identifier 10.1109/MTS.2018.2826058
Date of publication: 31 May 2018

h-index (a correlation of productivity and impact via paper and citation count), despite this index being primarily correlated with network centrality (i.e., popularity) [10] rather than a reliable measure of academic quality[6]; and journal impact factor, although similarly it has been argued that the figure alone is no guarantee of academic merit [2]. Both metrics are, of course, open to manipulation, and impact factor can even induce a state of scientific delusion: if some researchers are asked by peer reviewers for "one more experiment" before their paper can appear in one of the "top" journals, then they know what the outcome of that experiment *must* be[7] [4].

The expansion and its metrication creates a near perfect storm when it coincides with the "publish-or-perish" mentality and the imposition of national evaluation exercises, such as the U.K.'s Research Evaluation Framework (REF). For example, REF2014 was used to evaluate about 130U.K. universities employing approximately 200 000 academics, each of whom had to submit four publications they had produced in a six-year period. For the sake of argument, assume that these are all journal publications, supposing that non-journal publications are balanced out by four being the *minimum* number.

The gathering storm metastasises into the perfect one when one throws in the grand larceny masquerading as a public good otherwise known as Open Access. Multiply the number of academics by the number of their papers and the average fee charged for open access, and the Fermi equation/back-of-an-envelope calculation reveals a huge number — and this is in the U.K. alone. And this money is paid to publish the work of people that have already been paid to produce it…

It is at this point that the technological and business models underpinning the transition from print to electronic format in academic publishing find themselves perfectly positioned to exploit it. While many publishers do work very effectively with their journals, some publishing houses have used their historically-acquired position as guarantors of scientific quality and neutrality together with the new technology to create a *platform economy* [8], [12]. Moreover, the raw material is provided for free; the labor to convert the raw material into finished product is provided for free (i.e., editors, peer reviewers, etc.); distribution, advertising and promotion are provided for free; and even the growth of the market is provided for free — which is precisely where this editorial started, with the expansion of higher education.

In a platform economy, this exploitation is precisely what can be expected when both the ownership of the means of production *and* the means of coordination are privately owned. The only "winner" in the fallout from where technological advancement underpinning the transition in academic publishing clashes with societal changes is the enterprise.

This is not to suggest that centralization is the preferred alternative: the system of editorial and peer-review, for all its limitations,[8] has been crucial to the academic community, both in access control and quality control, as well as in self-governance and self-correction. The vast sums of money spent on open access should instead be invested in non-profit NGOs like the IEEE to foster the development of completely decentralized knowledge commons [5], [6] — using the same technology to create an open, democratic *platform community*, not a platform economy that enables some enterprises to accomplish their objective of maximizing profit by what some might argue is a form of legalized profiteering.

Author Information

Jeremy Pitt is Professor of Intelligent and Self-Organizing Systems at Imperial College London, U.K. Email: j.pitt@imperial.ac.uk.

References

[1] I. Asimov, *Foundation.* Gnome, 1951.
[2] A. Kurmis, "Understanding the limitations of the journal impact factor," *J. Bone Joint Surg. Am.*, vol. 85-A, no. 12, pp. 2449-2454, Dec. 2003.
[3] E. López, *The Pursuit of Justice: Law and Economics of Legal Institutions*. Palgrave MacMillan, 2010.
[4] R. Harris, "Perspective: Publish and perish," *Issues in Science and Technology*, vol. 33, no. 4, 2017.
[5] C. Hess and E. Ostrom, *Understanding Knowledge as a Commons*. Cambridge, MA: M.I.T. Press, 2006.
[6] S. Macbeth and J. Pitt, "Self-organising management of user-generated data and knowledge," *Knowledge Engineering Rev.*, vol. 30, no. 3, pp. 237–264, 2015.
[7] N. Oreskes and E. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming.* Bloomsbury, 2010.
[8] G. Parker, M. Van Alstyne, and S. Choudary, *Platform Revolution: How Networked Markets are Transforming the Economy — and How to Make Them Work for You.* Norton, 2016.
[9] J. Pitt and A. Nowak, "The reinvention of social capital for socio-technical systems," *IEEE Technology & Society Mag.*, vol. 33, no. 1, pp. 27–33, 2014.
[10] E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, and F. Schweitzer, "Predicting scientific success based on coauthorship networks," *CoRR*, abs/1402.7268, 2014.
[11] C. Tarvis and E. Aronson, *Mistakes Were Made, But Not By Me: Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts.* Mariner, 2015.
[12] M. Ulieru, "Blockchain and the real sharing economy: 'Uberisation' demystified," 2016; https://www.linkedin.com/pulse/blockchain-real-sharing-economy-uberisation-dr-mihaela-ulieru.

[6] Like Asimov's Psychohistory [1], *h*-index works passably well as an indicator, provided the subjects don't know they are being assessed by it. Otherwise, it has a sort of quantum effect, whereby taking a measurement of a system affects that system. It is well known that people don't just comply with rules, they react to incentives implied by the rules [3].

[7] Hence the equation: *career_pressure + confirmation_bias* = 1/*scientific_method*. For more on the issue of confirmation bias, see [11].

[8] It has impersonations too: the unscrupulous use of the scientific method (founded on doubt) against itself, by muddying the public understanding of a series of health and environmental issues, is charted in [7].

Book Reviews

Michael L. Black

Digital Research Confidential

Digital Research Confidential: The Secrets of Studying Behavior Online. By Eszter Hargittai and Christian Sandvig. M.I.T. Press, 2015.

Interest from academics in the humanities and social sciences in studying the cultural dimensions of computing can be traced back at least as far as the early 1980s. As personal computers became increasingly common in homes, offices, schools, and universities, scholars like Jay David Bolter, Lucy Suchmann, and Sherry Turkle began the work of adapting the research methodologies of their respective disciplines to study computing in these new contexts. Personal computing devices and the myriad of cultural activities we juggle through them have since become more and more complicated; however, at the same time user-friendly approaches to design encourage us to take this complexity for granted. Indeed, as Eszter Hargittai and Christian Sandvig note in their introduction to *Digital Research Confidential: The Secrets of Studying Behavior Online*, there is now often little institutional reward for writing about how the broad integration of computing affects academic research about culture and behavior. In this respect, their collection serves as an effective argument for the value of sustaining in-depth conversations on the effects that constantly changing technological conditions have on research methods. It is also an engaging introduction to the wide range of research in these fields being conducted digitally.

Not to be confused with the "digital humanities" — a field largely devoted to studying and preserving pre-twentieth century texts due to the copyright's influence over digitization efforts — the "digital media studies" represented in Hargittai and Sandvig's collection focuses on contemporary and often digitally created cultural activities. This interdisciplinary research area involves scholars from a variety of disciplines including, but not limited to, literature and language, writing studies, history, anthropology, sociology, communications, information science, and computer science. Representing every methodological approach in this incredibly diverse field in a single volume would be an impossible task. Hargittai and Sandvig have wisely chosen to limit their collection by focusing primarily on scholars whose work involves Web 2.0 technologies. This decision gives each of the essays in the book some common ground while still allowing the collection to highlight the breadth of subjects covered by digital media studies. The collection discusses projects involving the Internet Archive's Wayback Machine, YouTube, Twitter, computer-aided drafting software, Amazon's Mechanical Turk system, Wikipedia, Flickr, homemade webcrawlers, and Second Life.

Hargittai and Sandvig's excellent introduction is able to synthesize the wide-ranging research trends in this field around a tension between digital media as instrumentation and digital media as object of study. Digital research in the humanities and social sciences is often framed in one of two ways: either online platforms that offer "a new kind of microscope" allowing us to understand an area of offline behavior we are already familiar with in a new way, or online social activity sharing enough similarities to offline behavior that we can comfortably transfer our existing assumptions about how humans relate to one another into these new contexts. Both of these approaches assume that theory and method can be discussed separately. Most research leveraging these approaches, in other words, assumes either that digital tools add to — but don't disrupt our — core understandings of cultural and behavior or that online activities function as relatively seamless extensions of offline behavior. Yet as Megan Sapnar Ankerson concludes in her essay, each project in the collection involves a moment when a researcher "could not help but notice the ways (their) entire engagement with (their) project was thoroughly organized through software" (p. 47). While the widespread acceptance of particular methods in the humanities and social sciences allows for basic assumptions about culture and society shaping them to be taken for granted, digital tools are often designed with other needs in mind. Conducting social research

Digital Object Identifier 10.1109/MTS.2018.2839218
Date of publication: 31 May 2018

in digital contexts, which is to say through software, requires us to revisit and re-evaluate them against the affordances and constraints of the tools we use. *Digital Research Confidential* thus makes a strong argument for both the practical and analytical value of software and data carpentry in the humanities and social sciences.

The ten essays that follow the introductory chapter can be roughly divided into four categories. The first two essays following the introduction explore how database construction shapes archival research practices. Megan Sapnar Ankerson shares her experience working with the Internet Archive's Wayback Machine. Initially taking the project's slogan to "surf the web as it was" at face value, Ankerson eventually realized that some sites were reproduced in a form that had never existed in the first place due to the way that the project's webcrawler tried to fill in gaps in content by selecting temporarily adjacent versions of files with the same name. Turning to debates taking place among web developers in the late 1990s, Ankerson is eventually able to connect the patterns of missing content resulting from the Internet Archive's retrieval algorithm to discussions about the norms of "good" web design. She concludes by observing that the Wayback Machine's algorithms are themselves a part of these debates in the sense that what they capture, recreate, or wholly ignore reflect beliefs held by its creators about what the web would become. In the chapter that follows, Virág Molnár and Aron Hsiao discuss a project on tracing the evolution of flash mobs using recordings published on YouTube. In addition to sharing some of their results, Molnár and Hsiao discuss how they documented their retrieval process in the interest of reproducibility, both so they could update their work later but also so that others might have a model to follow in similar projects. Yet when revisiting their results after the initial capture period, they discovered that YouTube's search interface had been altered, leaving them unable to reproduce their procedures. Like Ankerson, Molnár and Hsiao's essay demonstrates how digital research methodologies are influenced even before a project begins by technical decisions that researchers ultimately have no control over. While neither essay attempts to dissuade researchers from using public databases, they both point to a need to address the fact that decisions made in the name of efficiency or usability often have consequences for researchers that are not anticipated by software designers.

There is now often little institutional reward for writing about how the broad integration of computing affects academic research about culture and behavior.

Three essays in the collection discuss ethnographic methods, each demonstrating that close observation of the social use of technology requires creative adaptation to the various ways that technology is incorporated into our personal and working lives. The first, by danah boyd, discusses her work in documenting the roles that social media play in the lives of teenagers; however, boyd's essay rarely mentions technology. As she explains, her interviewing techniques reflect the ways that technology flitters in and out of their lives. While the Internet remains a significant conduit for teenage social behavior, she notes that it is rarely a focus of the narratives they tell about their lives. For anyone already familiar with boyd's work, this observation is not new, but the chapter offers important insights as to how she learned to adapt her ethnographic interest in computing to study a group that does not spend much time thinking about it. Technology also appears, at least initially, to exist at the margins of Paul M. Leonardi's essay on observing automobile engineers. Leonardi recounts his time conducting field research within a private company, offering a wealth of practical advice for every step of the process from explaining the idea of software as data in an Institutional Review Board proposal, to gaining the trust of management while on site. More importantly for the aims of Hargittai and Sandvig, Leonardi's work demonstrates the importance of incorporating a technological awareness into his observational method, as Leonardi is able to discover many of the disagreements he documents between engineers result from the way their software resolves conflicts when trying to integrate components designed by different groups before an impact test. Finally, Amy Bruckman, Kurt Luther, and Casey Fiesler's essay examines the role of Institutional Review Boards in universities, placing many of the same ideas raised in these two earlier essays into a historical context, showing how the social norms of online environments often diverge from the privacy and protection needs of offline social contexts.

Crowdsourcing has already proved effective in a number of research contexts, but as a pair of essays here show, the patterns of engagement it affords may conflict with some best practices of survey design and academic division of labor. Eric Gilbert and Karrie Karahalios' essay on developing a Twitter interface based on theories of "social ties" explores the difficult task of balancing the expectations of social science research against the feedback from user testing. Despite succeeding from a technological perspective, a question of representativeness lingers over the project, as start-up style beta testing is not necessarily compatible with norms of sample selection in the social sciences. Following their essay, Aaron Shaw's narrative of incorporating crowdsourcing into a large-scale content analysis project discusses the problem of user error in research software and expands into a broader reflection on how academic labor is valued. But this decision comes with its own time-consuming problems. Whereas professionally trained research assistants often disagree over interpretation, Shaw notes that disagreements in crowd-sourcing are often also the result of usability issues in the project's software interfaces and documentation. Shaw's account of addressing miscommunications between himself and software engineers on the importance of collecting data in specific ways will ring true to anyone who has participated in complex collaborations and is a must read for anyone considering one, regardless of their role.

The collection also includes a trio of essays on doing social research using "big data" retrieved from online sources. In their conversationally styled essay, Michelle Shumate and Matthew S. Weber offer two complementary, detailed defenses of software carpentry through a discussion of programming web-crawling tools. Their essay offers a look at the same issues raised by those before them in the collection but from the other side of the interface. While the labor of programming may not yet be institutionally rewarded in the humanities and social sciences, Shumate and Weber offer a strong argument for the ways that it can enrich those products of research that are. Brent Hecht and Darren Gergle extend discussion of this topic by examining how disparate data sources often make very different assumptions about metadata and its presentation within a database. Their essay raises important questions about the bias of certainty present in research involving big data. Similarly, Brooke Foucault Welles' discusses how larger data sets can actually create a greater degree of uncertainty by examining how seemingly familiar concepts like "friendship" are understood via comparatively narrow observational contexts. In addition to offering detailed looks at the labor behind large-scale data collection and preparation, these three essays show that close involvement in the algorithmic processes performing it will lead to increased confidence in the theories derived from it.

Digital Research Confidential's breadth, in short, is ambitious and representative of the way that digital media research in the humanities and social sciences is often peppered across academic departments on most campuses rather than centralized under a single umbrella. Readers will likely find that not all subjects or methods offer insights applicable to their own research. For instance, someone interested in combing through digital archives may not find the lengthy discussions on gaining trust during interviews as considerations of how databases prejudice our sense of historical memory.

> **Each project involves a moment when a researcher could not help but notice the ways their engagement with their project was thoroughly organized through software.**

If Hargittai and Sandvig have aimed the book at fostering conversations among seasoned practitioners in the fields represented about the long-term effects of recent developments in Internet-based technologies on their home disciplines, then in that regard it succeeds wonderfully even if every reader will take away something different from it. Nonetheless, each chapter is generally written in a manner beneficial to specialists and non-specialists alike, making the collection also an excellent choice for an upper-level classroom, scholars already wrestling with similar problems themselves, or technological professionals looking to better understand the computationally-driven research of their humanist and/or social scientist collaborators.

Reviewer Information

Michael L. Black is an Assistant Professor of English as UMass Lowell, Lowell, MA. His current book project examines the influence of market pressures and technical discourses on popular conceptions of software usability and personal computing.

Jenna P. Carpenter

Programmed Inequality

Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing.
By Marie Hicks. London, U.K.: M.I.T. Press, 2017, 239 pages.

Our awareness and understanding of the key role that women played in codebreaking and computing efforts during World War II has grown significantly over the last decade. While some stories, like that of pioneer computer programmer Grace Murray Hopper, have been known for some time, a more extensive story of women's impact has been painted in the last five to ten years by documentaries like *Top Secret Rosies: The Female "Computers" of WWI* [1], popular movies such as *The Imitation Game* [2], and books, including Margot Shetterly's recent *Hidden Figures* [3]. Marie Hicks' book, *Programmed Inequality* adds to this narrative by telling the story of the critical role that women "computers" played in Britain from World War II through the 1970s. However, Hicks seeks to go farther, recording not only the impact of women on the rise (and fall) of the computing industry in the U.K., but showing how the fortunes (or more accurately, the lack thereof) of women in the computing workforce were intertwined with Britain's inability to capitalize on the worldwide lead they enjoyed in computing at the end of WWII. Hicks also argues that this story is relevant today, given burgeoning U.S. computing workforce needs, positioned against the shortage of qualified computing workers. Indeed, like Britain, the U.S. worker shortage is exacerbated by the outmigration of women from the computing workforce that started at the end of WWII and accelerated in the early 2000s with the rise of the computing gamer culture [4].

> **The term "Computers" was originally coined to describe women who did mathematical computations by hand.**

How does Hicks's story go? Starting in the late 1800s, human "computers" (the term was originally coined to describe *women who did mathematical computations by hand*), were employed to support weather applications and astronomical research. Their calculations were done by hand and, later, with the aid of a series of desktop (electro)-mechanical machines (the predecessors of what we today think of as "computers"). Computing was actually viewed as a viable career path in the late 1800 and early 1900s for young women who showed mathematical talent. Around the 1920s in the U.S., the number of women entering college had grown to the point that people feared there would soon be too few slots for men, so a quota system was implemented to keep women out of fields like mathematics and science. But computing with the aid of desktop machines, even large floor-standing models like the IBM tabulator, were viewed as "secretarial work" — low-skilled, rote, minimally-valued. Therefore computing continued to be considered suitable work for women into the 1930s. Even the manufacture of computing machines at IBM in Britain was dominated by female employees, so much so that IBM measured its production in the U.K. in "girl hours," instead of the more common "man hours" until the 1960's. Layered on all of this was the gendered expectation that women's primary goal in life was to marry and have children. Working was a temporary diversion for young women in their late teens, early twenties. When women married, they were expected to drop out of the workforce. After all, their

Digital Object Identifier 10.1109/MTS.2018.2839219
Date of publication: 31 May 2018

husbands should be able to financially support them. A workforce culture built on these assumptions meant that women's pay was kept low and opportunities for promotion or positions in management were all but non-existent.

World War II dramatically challenged the artificial rules about the suitability of women in the workplace. With vast sectors of male workers siphoned off to the military by the draft, countries like Britain had no choice but to open the doors of the workplace to women. Hicks notes that by 1942, Britain was relying so heavily on women in the workforce that it was forced to open up all jobs to women, including the heretofore male bastions of engineering and welding. While employers surprisingly found women perfectly capable of doing such work, they still treated only the most talented women equal to their average male counterpart. Yet industries in the U.K. went so far as to reorganize how work was done and how training took place, in order to make it easier for women to participate. After all, as Hicks points out, 1.1 million women (80% of single women 41% of wives and widows, 13% of mothers with children under 14) were hired in 1942 in the British armed forces and munitions industries alone. The numbers only increased as the war dragged on.

Over 10 000 women worked at the famous Bletchley Park, mostly young, single, white, and middle class. These women worked on decoding German communications, much of it using British Colossus computers. The Colossus preceded the creation of the U.S. ENIAC computer at the University of Pennsylvania, which means Britain led the world in the development of the modern computer. Despite the centrality of their work to the war efforts, women at Bletchley Park were not promoted into positions that reflected their newfound skills, the type of work they performed, or their future potential. Even severe wartime manpower shortages were not enough to completely override the gendered work culture in the U.K. Underage and inexperienced teenage boys, for example, were recruited and trained to be maintenance engineers for the computers at Bletchley, bypassing the older, experienced women. Indeed, the overarching message of Hicks' book is the length to which Britain went to maintain its gendered work rules and culture, no matter how damaging the consequences. Hicks illustrates how these same gendered rules were one of the key sources of the country's computing workforce woes in the latter half of the 20th century.

The complete secrecy — long after hostilities ended — of wartime activities at Bletchley Park and similar installations in both Britain and the U.S. meant that the accomplishments and skills of women in the wartime computing effort were never acknowledged or made public. This fueled for many more decades the faux storyline that women were not interested in, nor capable of performing computing and related technical tasks. At the end of WWII, only several hundred of the thousands of women at Bletchley were allowed to transition over to the then-coveted government Civil Service computing jobs, the first of what would grow to be Britain's enormous postwar, largely female computing workforce. At odds with reality, however, the women's computing work inside the Civil Service system was regarded as low level and subordinate to the "real work," instead of squarely at the core, of postwar computing. The roles women were allowed to fill were seen as separate and beneath the work in which their male coworkers engaged. Consequently, the Civil Service system in Britain was, for decades, among the worst perpetuators of the unequal categorization of men's and women's computing work. Campaigns for equal pay and other opportunities for women arose time and again in postwar Britain, yet the country repeatedly managed to dodge any real change. Between the end of WWII and 1946, married women were prohibited outright from working in Civil Service jobs (due to the so-called "marriage bar"), even though other industries began to relax such rules since they had been ignored during the WWII. The 1946 Equal Pay Report removed the marriage bar, yet still assumed that the number of women in the workforce would be inconsequential and only allowed under certain restrictive circumstances, certainly not a key part the nation's postwar economy. Because women in Britain had few opportunities for advancement or interesting work after WWII, many who had worked during the war years left the workforce voluntarily after they married.

A quota system was implemented to keep women out of fields like mathematics and science.

After repeal of the marriage bar in 1946, the British Civil Service created a substandard system called "machine grades" for women's computing work, in a deliberate effort to limit women's pay and advancement opportunities. Women's work still largely consisted of working with machines, perpetually viewed as low-skilled secretarial work and therefore appropriate for women but beneath men. The "machine grades" succeeded in devaluing women's work, pay, and computing careers in the U.K. for decades. The creation of the "machine grades" also successfully convinced the British people that computing was a low-level occupation that required no real education or skills. By the time the British government realized that computing was revolutionizing how work was done in the latter part of the 20th century, it was too late. No one believed the new government propaganda that computing was a high-level job offering a lifetime of career opportunities.

After WWII, the British government took over about one-fifth of the country's private industries and continued to enforce wartime practices such as rationing. The socialist agenda of the British government during the post-WWII period did not encompass equal pay for women. Although this issue was revisited multiple times in the ensuing years, the perpetual argument was twofold. First women did not need to make as much as men (who had to support a family). After all, a single woman only had to take care of herself and a married woman had her husband's salary on which she could rely. The second argument against equal pay was that the government could not afford to pay women the same salaries that they paid men. Neither of these arguments was accurate. Many families needed the extra income that a working woman provided. In fact, Hicks points out that the government spent far more than needed to fund equal pay for women on social programs aimed at accomplishing the same results that equal pay would have achieved. The British Civil Service worked to implement major cost reductions during this time, but believed that keeping inflation in check could be accomplished by hiring cheaper women workers and by transferring much of their work from humans to computers. Therefore, Britain was vested in keeping women's pay artificially low as a matter of national economic policy and they devoted enormous resources over several decades toward that goal. Britain failed to foresee that their computing needs, in terms of both equipment and trained workers, would mushroom and consequently so would the financial resources required to maintain them. In the end, neither artificially depressing women's wages nor relying on computers as a cost reduction measure was a viable economic strategy.

"Machine grades" succeeded in devaluing women's work, pay, and computing careers in the U.K. for decades.

In the 1960s and 1970s the British government sought to rebuild its wartime dominance in the computing sector to restore its position as a global power. As noted above, this plan necessitated a gender power shift in computing, from viewing computing as low-wage, unskilled secretarial. or "machine grade" work performed by women, to seeing it as high pay and prestigious work performed by men. In Britain's effort to rebrand computing as "high value men's work" they booted out their qualified and experienced workforce of women and tried to replace them with inexperienced and largely uninterested men. The fact that the government's efforts here were a flop necessitated increasing government micromanagement of the private British computing industry. Their inability to attract a sufficiently large male computing workforce (and, of course, now that computing was important work, women could no longer be allowed to engage in it) meant that Britain's enormous Civil Service sector needed powerful supercomputers that could be "controlled" by a small number of high level, trained male executives. Note that this obsession with supercomputers took place while the rest of the world was focused on the emergence of the personal computer or PC. But the British government insisted on dictating the future direction of the U.K.'s computing industries, shoring them up with government funds and eventually forcing a merger into a single computing company capable of designing and building the colossal supercomputers to run the country. Hicks meticulously details the story of Britain's intertwined desire for world power status in computing with its economic woes as well as its stubborn adherence to a strongly gendered workplace. She also demonstrates why this approach caused Britain to fail, sealing the demise of both its reemergence as a world power and its dominance of worldwide computing. Instead of being a fix for Britain's economic and political power woes, computing and the changes it ushered in simply exacerbated the inherent problems embedded in both systems. Only in the late 1980s did the British government finally engage in efforts to specifically recruit women into formerly male computing jobs. They hoped to alleviate the decades-long computing worker shortages and outflow of even minimally trained men to the more lucrative private sector. Yet these efforts to re-engage women in the British computing workforce failed by the end of the 1980's and never recovered. This loss of what once was an enormous female Civil Service computing workforce finally succumbed to decades of intentional discrimination.

How is this story relevant today? Hicks rightly notes that the number of college-aged women pursuing a career in computing has plummeted in the last 15 years in the U.S. (4). But the problems start long before women get to college. By the time U.S. children are in second grade, research shows that they all know math is for boys and reading is for girls. By the time children reach high school, boys outnumber girls a stunning four-to-one among Advanced Placement or AP Computer Science test takers. In 2014 three U.S. states (Mississippi, Montana, and Wyoming) had zero girls take (not pass, just take) the AP Computer Science exam. Zero (4). Efforts to attract girls to computing starting in elementary school in the U.S. are numerous to be sure, from coding organizations girls like Girls Who Code and Black Girls Code to recognition efforts such as NCWIT Aspiration Awards programs to K12 robot competitions like FIRST and Vex Robotics. It will be some time before we can truly assess the long-term impact of such efforts. Programs to recruit and retain college women in computing-related majors have been around for decades, as well. There are some success stories at institutions such as Harvey Mudd College, where more than half of the computer science graduates in 2016 were women (5), but these stories are the exception, not the rule. Computer science programs in college rank among the lowest in terms of percentage of women majors, often hovering not far above the single digits. Like postwar Britain, the computing workforce needs in the U.S. are huge and growing rapidly, with an estimated 1 million more computing jobs than qualified applicants by 2020 (6). Combine all of this with the fact that women now outnumber men in college two-to-one (7), and it's not hard to see that U.S. is already riding a tsunami fueled by the shortage of qualified computing workers. It's just that the bulk of the water hasn't obliterated the shoreline ... yet. Given that the bachelor's degree college graduates of 2020 arrived on U.S. college campuses in fall of 2016, we will have to increasingly rely on alternate training, such as coding camps, two-year programs, retraining of older workers and, yes, successfully cracking our own gendered computing norms in the U.S., to survive. The country with the best trained and largest computing workforce will likely rule the world in the not-too-distant future. We daily see increasing threats and damage by hackers on critical industries and infrastructure dotting the daily news.

As a female professional who works hard to attract and retain women in STEM careers, I certainly hope that we can pull it out in the end. But I am afraid that it is already too late and Hicks' warning will simply go unheeded. Because, like postwar Britain, our cultural gender norms about who can and wants to do computing in the U.S. have actually grown more restrictive over the decades (just look back at those AP Computer Science test takers!), hurling us in the opposite direction of where we need to go, driving women away from computing fields at the very time we desperately need to be drawing throngs of women toward careers in computing with open arms.

In Britain's effort to rebrand computing as "high value men's work" they booted out their qualified and experienced workforce of women and tried to replace them with inexperienced and largely uninterested men.

Reviewer Information

Jenna Carpenter is Founding Dean and Professor of Engineering at Campbell University, 143 Main St, Buies Creek, NC 27506. Email: carpenter@campbell.edu.

References

(1) L. Erickson, *Top Secret Rosies: The Female "Computers" of WII*. PBS (Film), 2010.

(2) M. Tyldum, *The Imitation Game*. The Weinstein Company (Film), 2014.

(3) M. Shetterly, *Hidden Figures: The American Dream and the Untold Story of the Black Women Who Helped Win the Space Race*. New York, NY: William Morrow, 2016.

(4) "The current state of women in computer science," *computerscience.org*; http://www.computerscience.org/resources/women-in-computer-science/, accessed July 9, 2017.

(5) M. O'Sullivan, "Women thriving in computer science at California college," *Silicon Valley & Technology, Voice of America*, Feb. 5, 2017.

(6) J. Swartz, "Businesses say they just can't find the right tech workers," *USA Today*, May 9, 2017.

(7) M. Whaley, Men saying "no thanks" to college," *Denver Post*, June 5, 2017.

Renée M. Blackburn

Girls Coming to Tech

Girls Coming to Tech! A History of American Engineering Education for Women.
By Amy Sue Bix. M.I.T. Press, 2014.

Histories of Science and Technology in the twentieth century often deal with important figures, almost certainly male, and their contributions to their field and to society. These histories detail the significance of their education and the ways in which they used their skills to start major companies, create new technologies, or contribute to patriotic efforts in wartime. In *Girls Coming to Tech!: A History of American Engineering Education for Women*, Amy Sue Bix expands on the growing literature of women in STEM. Specifically, the book explores engineering education through the lens of gender from the late 1800s through most of the twentieth century.

The text comprises seven chapters, each highlighting the measures taken by institutions and individuals to eventually reach engineering coeducation. Chapter 1 focuses on the few women who first entered institutions in the late nineteenth and early twentieth centuries. Chapters 2 and 3 indicate the important role that World War II played in shifting social and institutional norms around women's engineering education. Not only did universities accept more women into engineering, so did the federal government and industry, particularly through partnerships that blurred gender boundaries temporarily for the war effort. Chapters four through six offer case studies of post-World War II women's engineering life at Georgia Tech, Caltech, and M.I.T., respectively. Each university approached coeducation from a different angle, with varying cultural norms attached to women's femininity and how those views fit into the heavily masculine traditions already in place at each institution. And finally, chapter seven details the ways in which campus culture and views around coeducation changed in the late twentieth century.

Bix, Professor of History at Iowa State University, uses the larger narratives of American history, mostly in the twentieth century, to explain the way in which engineering education for women was viewed and decided by universities, industry, and the federal government. In one of the more prominent examples, Bix, at various points, details the impact of World War II on expanding women's entrance into engineering programs and women's general acceptance at technical institutions and in male-dominated environments. She describes the stereotypes that plagued women who wished to pursue engineering: they only wanted to find a husband, they would quit work once they were married or had children, they were not smart enough to grasp the science or math, or they were only suited to engineering programs that already possessed feminine qualities.

Bix expertly draws out the threads of masculinity and femininity that guided many of the actions taken by actors throughout the story. Not only were engineering programs averse to women entering the "masculine" field of engineering, but when they did, programs often sought to play on what they saw as women's inherent feminine characteristics. In one example, Purdue created a new program called "Housing" for women after World War II, which married the already feminine Home Economics with Engineering, "based on assumptions that female students were natural authorities on the home." This worked both ways, of course, as Purdue hoped not only to appeal to women interested in engineering and who did not wish to enter the engineering school, but also appealed to a growing number of men who were interested in majoring in home economics.

Throughout the text, Bix also draws on the "invasion" metaphor, wherein male students, faculty, and administration saw female students as "invading" campuses when they first arrived. This metaphor is particularly effective to convey the way

Digital Object Identifier 10.1109/MTS.2018.2839220
Date of publication: 31 May 2018

both men and women felt about women's place on college campuses. Male students often worried, unnecessarily, that the introduction of women to classrooms and campuses would both diminish the standards set for their education and distract them from their work. At the same time, they held a double standard against female students for being both "distracting" and yet undateable. On the other hand, women felt isolated and alone. They were often singled out and treated inappropriately — with preferential treatment, harassed by classmates and faculty, or ignored or spoken over in meetings or group settings.

One of the lessons of this text is that women's entrance and engagement with engineering was not a linear progression. Some engineering programs, like those at RPI, Purdue, and Columbia, expanded their programs and allowed women to enter engineering programs or universities at various points in time before World War II. At the same time, these universities also guided women toward gendered course structures, offering courses in subjects such as "domestic arts," which they assumed women wanted and needed. University administrations' motivations for admitting and educating women students remained firmly entrenched in gendered notions of women's and men's places in society, and even college educated women were assumed to be in need of domestic skills only. The few women who did enter and graduate, however, were often still met with the same "masculine" mentality that they experienced while at university.

During World War II, a rush of women entered into engineering programs across the United States, compared to previous years. Though many of the women who had entered wartime programs such as the Curtiss-Wright Cadettes, did not stick with engineering after the war ended, they created a more hospitable climate for women seeking engineering degrees. Engineering programs during World War II helped to blur the strict gender lines that previously kept women from easily pursuing engineering degrees, but did not allow women full access to the same education system that their male engineering counterparts experienced. The double standard of engineering education continued when programs were specifically designed for women.

Bix devotes three chapters to specific case study institutions and their varied reactions to coeducation. Chapter 4, "Coeducation via Lawsuit: Georgia Tech," and Chapter 5, "Coeducation for Social Life: Caltech," approach coeducation from the students' point of view. Male students recognized the need for coeducation to create a better social world at Caltech. However, the male students and faculty treated the incoming women as outsiders and objectified them based on their appearance and perceived abilities. The third case study institution in Chapter 6, "A Special Case: Women at M.I.T.," was coeducational in the nineteenth century, though women's large-scale enrollment did not occur until the mid-twentieth century. However, while other institutions were debating whether to allow women in, M.I.T. debated whether to begin excluding women. In the mid-1950s, M.I.T. commissioned a committee on coeducation, and many on the committee argued for making M.I.T. male-only as they saw some women students in their courses struggling to succeed. But the larger problems that women dealt with at M.I.T., such as lack of adequate housing, added to their difficult experiences and continued to aid the cause that women did not belong in the masculine technical world of M.I.T..

In all three cases, the women at tech schools faced double standards. They were welcomed through new admissions initiatives targeting them, but upon entering were often treated as sexual objects or intellectually inferior by their male peers and faculty and provided with inferior services like housing, making their transition to university life more difficult. They hoped that once they reached a "critical mass" of women students, the invasion rhetoric and resistance to their presence would fade as women gained more respect.

Beyond the three major case studies found in Chapters 4, 5, and 6, Bix provides multiple examples from universities across the United States, including of schools like Cornell, University of Michigan, University of Minnesota, and University of Colorado. Her book is well-researched and her conclusions reverberate beyond the university and into the role that women play in the workplace, especially in technical positions, and in the ways that science, math, and engineering are taught to girls at young ages. This book is more than a history of women's struggles in gaining a university-level engineering education, it is also a history of the changing societal notions of what technical education is and should be and how that education should be accessed, taught, and used.

Reviewer Information

Renée Blackburn is a Ph.D. candidate in the History, Anthropology, and Science, Technology and Society (HASTS) program at M.I.T.. Her dissertation work focuses on traffic safety and technology policy in the second half of the twentieth century, exploring changing ideas of gender, freedom, and responsibility in American culture. Email: rmblack@mit.edu.

Shoshana Eilon

That Dragon, Cancer

Video Game as Art Form

THAT DRAGON, CANCER/NUMINOUS GAMES

In 2012, when Ryan and Amy Green learned that their baby son Joel's rare cancer was terminal, they were devastated. Searching for a way to explore his feelings, Ryan, an indie video game developer, found solace in the most appropriate creative outlet he knew: making a video game.

Ryan Green created a video game called "That Dragon, Cancer," a game that is at once a poetic exploration of a father's relationship with his son, an interactive painting, and a vivid window into the mind of grieving parents.

Green also recruited his wife and sons into the process of documenting their daily lives for a film, *Thank you for Playing,* about the video game's development. In creating the documentary and the game, Ryan had to decide where to draw the line in sharing his family's experiences of raising a dying child. From having his sons reenact difficult conversations, to recording Joel's giggle, to painstakingly photographing every detail of the hospital, Ryan's life became consumed by the complicated process of creating a digital world that mirrors his own, even as he continued to care for his son.

Developed by parents of a terminally ill child, "That Dragon, Cancer" facilitates emotional connection and spiritual awakening.

Combining footage from both Ryan's real and animated worlds, *Thank You for Playing* examines how we process grief through technology in the twenty-first century, and the implications of documenting profound human experiences in a new artistic medium: the video game.

"That Dragon, Cancer," the video game, was developed by Ryan and Amy Green and Josh Larson along with five others at their studio, Numinous Games. It was released in January 2016. As the Numinous team describe it: "'That Dragon, Cancer' is a video game developer's love letter to his son; an immersive, narrative video game to inspire love for others; a memorial for hundreds who have fought cancer. It is a poetic and playful interactive retelling of Joel Green's 4-year fight against cancer, and an autobiographical memoir of how parents Ryan and Amy embrace hope in the face of death."

Digital Object Identifier 10.1109/MTS.2018.2839221
Date of publication: 31 May 2018

The directors of the "Thank You for Playing" documentary were David Osit and Malika Zouhali-Worrall. In a statement, they noted: "Ryan and Amy's video game, 'That Dragon, Cancer' comes at a time when video games and interactive media are emerging as a wildly innovative art form. And yet simultaneously, society is questioning our dependence on technology more than ever: it seems to be bringing us at once closer together and yet further apart. [As Directors,] we are fascinated by this tension, which is why we set out to make this film.

"From the moment we first heard about "That Dragon, Cancer," we immediately wanted to know more about why Ryan and Amy had chosen a video game — a medium so often associated with explosions and violence — to convey one of the most emotional and spiritually-challenging experiences a family can go through. Once we saw for ourselves how many people were profoundly moved by the game, and how playing it often facilitated more, rather than less, social interaction, we were hooked and knew we had to keep following this story. The fact that a video game was capable of awakening this sort of empathy to allow players to join Ryan and Amy on their journey astounded us, and we soon realized that Ryan isn't only a video game developer, he's also an artist — and programming is his paintbrush.

"Thank You for Playing explores the very personal experiences of a family battling cancer, and the beauty and hope that can be found in the artistic process, while also examining the age-old question of where the boundaries lie in representing difficult emotional experiences in art. Ultimately, we hope [*Thank You for Playing*] the film will challenge people to reexamine their own assumptions about bereavement, technology, video games, and art."

The "That Dragon, Cancer" game can be found at thatdragoncancer.com. *Thank You for Playing* appeared at the Tribeca Film Festival in 2015, and is available for download at http://www.thankyouforplayingfilm.com/watch/.

Author Information

Shoshana Eilon is Director of Distribution at Film Platform, London, U.K. Email: shoshana@filmplatform.net.

PRESIDENT'S MESSAGE

(continued from page 1)

a previous offer to volunteer, please accept my sincere apologies and contact me again so I can assist you.

Call for Donations, Gifts, and Bequests

SSIT's 2018 fundraising campaign is focused on securing the level of resources required to scale activities over the coming years. Funds will be invested in further strengthening and expanding volunteer activities. Options to financially support SSIT volunteer activities include:

- Donate to SSIT online https://ieeefoundation.org/ieee_ssit.
- Mail a check payable to the "IEEE Foundation – SSIT Fund" to: IEEE Foundation, 445 Hoes Lane, Piscataway, NJ 08854, U.S.A.
- Asking your employer to match your personal donation.
- Donate in honor or memory of someone who has touched your life or others.
- Direct a gift to the "IEEE Foundation – SSIT Fund" from your donor advised fund, foundation or family office.
- Providing a legacy Remember SSIT in your will.

Author Information

Paul M. Cunningham, 2017–2018 IEEE-SSIT President, is President & CEO, IIMC (Ireland); Director, IST-Africa Institute (www.IST-Africa.org); Adjunct/Visiting Professor, International University of Management (Namibia); and Visiting Senior Fellow, Wrexham Glyndŵr University (Wales). He is 2018 Chair, IEEE Humanitarian Activities Committee and serves on the IEEE Global Public Policy Committee. Email: pcunningham@ieee.org.

Rebekah Shirley Daniel Kammen

Mundane Is the New Radical

The Resurgence of Energy Megaprojects and Implications for the Global South

Katse Dam in Lesotho, Africa.

Energy infrastructure is critical to the future of any rapidly emerging economy. Unprecedented rates of growth in the global south have quickly raised the stakes for finding plentiful, low-cost energy technology options to keep pace with development needs. This demand has been a significant factor but is not the only one driving a global resurgence in the deployment of large energy infrastructure, and in particular, the hydroelectric dam. Nevertheless, the increasing number of dam projects deployed in developing countries over the last two decades that perform poorly regarding their economy, the environment, human rights, inequality and wealth distribution, as well as public support, all illustrate a seeming disconnect between planners, stakeholders, and our technological energy solutions of choice.

The literature generally focuses on a techno-managerial assessment of large-scale energy projects, highlighting issues of technical and economic performance, environmental risk, and the impacts of social displacement. Beyond economistic and technocratic analyses of impact and mitigation, we argue that truly comprehensive energy project assessment should consider the contemporary and historical global contexts within which such developments are embedded. That is, we argue for examining the processes that give rise to energy projects, alongside consequences thereof. Such an assessment shows that balancing the need for large energy infrastructure with local and contextualized solutions is a major challenge that, more than technological dynamics, may be a challenge of cultural dynamics. We posit that addressing such seemingly mundane issues is the radical solution needed for sustainable infrastructure development, by exploring global drivers of the dam resurgence and discussing implications for policy.

Digital Object Identifier 10.1109/MTS.2018.2826076
Date of publication: 31 May 2018

Global Drivers of the Large-Scale Energy Infrastructure Resurgence

The Great Economic "Convergence"

The economic separation of early industrializers from the rest of the world during the Industrial Revolution, often termed "the great divergence," has characterized our global political and economic hegemony for the past two centuries [1]–[3]. But now, a historic change is taking place. A "great convergence" is underway as less developed countries quickly adopt the technology, competence, and policies that formerly propelled the developed world [4]–[6]. United Nation's Human Development data shows that for the first time in over 150years the combined output of today's most populous emerging markets — China, India, and Brazil — is equal to the combined GDP of all the major industrial powers of the north — Canada, France, Germany, Italy, the United Kingdom, and the United States — representing a major rebalancing of global economic power.

In fact, it is projected that China, India and Brazil alone will make up over 40% of global GDP by 2050 [7, p. 13], and the "convergence" is far beyond these three [8]. Countries such as Mexico, Bangladesh, Tanzania, and Yemen and at least forty others have registered significant growth this decade and other breakout nations such as Afghanistan and Pakistan had some of the fastest growth rates in the world over the past ten years. On average, non-oil, non-small developing countries have seen GDP per capita increasing at a rate of 3% per year since the 1990s. Today, the South produces about half of world economic output, up from a third in 1990 [7, p. 13]. While there have been periods of rapid growth for individual countries in the past, seldom in the last 50 years have we seen episodes where so many poor countries have simultaneously done well as in the decade preceding the recent Global Financial Crisis [9].

The evidence is clear, says the UN Human Development Report 2013, that "the rise of the South is unprecedented in its speed and scale. Never in history have the living conditions and prospects of so many people changed so dramatically and so fast. This change represents a global rebalancing far greater than that experienced during the Industrial Revolution. The Industrial Revolution was a story of perhaps 100 million people, but this is a story about billions of people" [7, p. 11].

This change in economic dynamics over the past decade is due in part to the differing experiences of Northern and Southern countries during and after the Global Financial Crisis of the 21st century. In the past, Northern countries served as the major importers of goods from Southern countries, such that as Northern economies grew or receded and as demand increased or decreased it would have a trickle-down effect on the export economy of less developed countries. The 21st century recession that resulted has largely upended this relationship.

It is argued that in developed countries the crisis stemmed from, in part, a constriction of credit flow, which followed the burst of the housing and oil price bubbles caused by excessively low interest rate policies from financial institutions [10]. Initially emerging economies "dodged the housing crisis that froze credit markets in the United States and Europe and that threw the rich world into the worst downturn since the 1930s. They never had to bail out their banks or endure the high unemployment and stagnant growth that historically follow financial crises" [11]. While the reduced spending and reduced demand from markets in advanced countries did eventually have impact on less developed countries, they were able to keep growing in the aftermath of the crisis, albeit more slowly, unlike most advanced economies which registered negative growth for many years after the crisis.

Economic growth alone does not automatically translate into human development progress. But Southern countries are not just tapping into global trade, they are also improving health, communication, and education services, which continue to support the growth experienced since the 2000s. This contrasts with contemporary policies adopted by many Northern countries which include austerity measures and cutting of social programs post-economic crisis. Experts say that it is this combination of policies, population growth and global economics that has allowed the middle class in the South to expand so rapidly [7]. In fact the UN projects that by 2030 more than 80% of the world's middle class will reside in developing countries and account for 70% of total consumption expenditure globally [7, p. 14].

Global South's Growing Middle and Energy Demand

With this unprecedented improvement in aggregate human development scores, we are now seeing an increasing demand for basic services across the globe. Improved water and sanitation access along with reliable energy services have become major Millennium and now Sustainable Development Goals (SDGs). The U.S. Energy Information Administration (EIA) projects that due to population growth, non-Organization for Economic Cooperation and Development (OECD) economies will account for more than half of the world's total increase in energy consumption until 2040, at which point they will account for two thirds of world total [12, p. 1]. In contrast, more mature energy-consuming and slower growing OECD countries will see total energy use increase only 18% by 2040.

This is compounded by the fact that energy consumption "per person" is also predicted to rise as developing countries grow not only bigger (more populous) but richer, as mentioned in the previous section. As middle-income groups in these countries grow larger, demands for improved standards of living, such as for better housing and sanitation, increase. As demands for housing, appliances, and transportation increase, energy capacity must also increase to produce food, infrastructure, goods, and services for both domestic and foreign markets, leading to higher per capita energy consumption. Whereas energy use per capita will remain flat in OECD countries over the next 30 years, EIA forecasts more than half the increase in global energy consumption will come from non-OECD countries across Asia, the Middle East, Africa, and Latin America in the same time period, even accounting for efficiency gains [12, p. 8].

We are seeing an increased focus on the need for electricity services in places where it has not been *as* major a human development focus before. Infrastructure has moved from being a "simple precondition for production and consumption to being at the very core of these activities" [13, p. 2]. This energy "pivot" to the South has given rise to a surge in large-scale energy infrastructure projects to facilitate industrial productivity and consumption [14].

Emerging Role of the Global South in Climate Change Mitigation

At the same time that energy demand grows sharply in the global south, there is also currently an increased global awareness of climate change and an international commitment to reducing emissions to limit temperature to under a 2 °C increase over pre-industrial levels. This was recently affirmed as the Paris Agreement was ratified by over 140 countries [15]. In the past, world leaders have argued that rich, industrialized countries created the global warming problem with their industrial emissions and should bear the larger brunt of emissions reduction — this has been a well-known sticking point in past climate negotiations [16], [17].

But climate experts and now even officials from developing nations are saying "there is no way that global warming can be kept below the international 2 °C goal without dramatic limits in future emissions from the developing nations [because] under a Business As Usual (BAU) scenario, most emission growth will come from the anticipated increase in fossil fuel use by developing nations" [18]. Experts find that approximately two-thirds of avoided emissions will have to come from the developing world to meet the collective goal, which means that new targets such as the "High Ambition Coalition" target of 1.5 °C, which while making very sound climate sense, poses particular challenges for developing nations [18].

Given the threat of global warming and the yet essential nature of electricity to development, low-emission energy solutions that supply massive amounts of power are in high demand [19]. This brings us to the hydro-electric power dam, our large energy infrastructure technology of focus.

Southern Investors and New Finance for Development Projects

Historically speaking, dams and hydroelectric infrastructure have always been on the international and national development agenda for modernization. Such projects were generally financed by international development cooperation agencies and multilateral development banks (MDBs). But the World Bank eventually came under strong fire for its lack of attention to the negative impacts of many of these projects, particularly regarding population displacement. The late 1990s were "characterized by escalating debates over large dams" [20] and fierce discussions over a number of high profile cases such as India's controversial Sardar Sarovar Dam.

Furthermore, cost overruns are typical and well-documented in hydropower finance. A recent Oxford study analyzed a sample of large dams built between 1934 and 2007 and found that three of every four dams suffer from cost overruns, one of every two dams had costs that exceeded benefits, and that the actual cost of dams is on average double their estimated costs [21].

Mounting international pressure arose against dams during this period. The World Bank was eventually forced to pull out of the Sardar Sarovar project after an independent review in 1993 [22]. The participation of MDBs in large-scale dam projects quickly subsided. At the World Bank alone investments in hydropower declined by 90% between 1992 and 2002 [23]. Consequently, there was a noted lull in international megadam funding during the 1990s.

Yet at the same time other events were brewing. In the aftermath of the 1997 Asian Financial Crisis, several Southern countries began developing new monetary arrangements for lending. Added to the South's growing financial reserves coming out of the 2009 Global Financial Crisis, this has transformed global financial architecture, such that the South has now become a major source of foreign direct investment (FDI). Most important is that this Southern investment is directed back to the South [23].

The World Bank, the International Finance Corporation (IFC), and the World Trade Organization (WTO) all acknowledge the dramatic uptake of South-South FDI. In fact it is projected that South-South trade will soon overtake trade between developed nations [24]. The combined value of FDI outflows from Brazil, Russia, India, China and South Africa (the "BRICS" countries) alone skyrocketed from U.S.$7 billion in 2000 to U.S.$126 billion in 2012, with nearly 58% being received by other developing countries. So though still a relatively small volume of total direct investment outflow, South-South FDI is growing at an annual rate of 21% [23].

Developing Asia is the largest recipient of FDI inflow, and accounted for nearly 30% of global FDI in 2013. China has strengthened its position as "one of the leading sources of FDI, and its outflows are expected to surpass inflows within two years." Flows to African countries have also increased significantly. Between 1992 and 2011, China's trade with Sub-Saharan Africa alone rose from U.S.$1 billion to more than U.S.$140 billion [7], [25]. Africa's FDI inflow increase is sustained in part by growing intra-African flow, from growing consumer markets. The share of investment projects originating from within Africa increased to 18% in 2013 from 10% in 2008 [26, p. 19]. This intra-regional investment front is led by Transnational Corporations (TNCs) from South Africa, Kenya, and Nigeria.

The rapid growth of Chinese outward FDI by both state-owned and private Chinese corporations was also catalyzed by deregulation. The Chinese government has been actively encouraging firms to invest overseas through its "Going Out" policy since 2000 [27]. Then in 2009 the Law Concerning the Control of Outward FDI by the Chinese Ministry of Commerce came into force, transferring authority to approve investment plans to local governments and greatly simplifying application criteria and process. In one year Chinese outward to inward FDI ratio jumped 10 percentage points and from 49% in 2009 to 55.8% in 2010 [28, p. 148]. As a result, "China significantly expanded its resources and energy availability base, in addition to gaining a foothold in the global manufacturing sector" [28, p. 147]. Many southern national and multilateral development banks, such as the Asian Infrastructure Development Bank (AIDB) also expanded global development financial flows, with banks able to craft their own lending policies as outward FDI became increasingly deregulated.

Thus national development banks and private investors from emerging economies such as China, Brazil, Thailand, and India have picked up the slack in international investment where MDBs like the World Bank left off [23], [29]–[31].

A Critical Culmination: The Large Dam Resurgence

All these conditions combined provide the ingredients for a great resurgence. Increasing energy demand in the global south is being partly driven by changes in the global economy and together with increasing focus on climate change mitigation commitments from the South act as a driver for low-emission technologies that deliver massive amounts of power — ostensibly in the form of projects such as the mega-dam. New investment opportunities for such projects have emerged from the south, filling the gap left by a northern MDBs financing downturn.

And indeed this is the boom we are seeing — globally, between 2005 and 2011, newly installed hydropower capacity outpaced new generation capacity from all other renewables combined, driven mostly by hydropower development in Asia, led by China, where — as discussed earlier — energy security has become a significant concern for sustaining its economic development [32], [33]. Already home to more than half the world's dams, China has built 850 more since 2000, scores of these since 2005. India has added 296 dams since 2000 and together countries like Brazil and Peru in the Amazonian basin have built or are planning over 400 new dams [34]. Indeed, new and resumed construction of mega-dams is underway across the global south, from Latin America to Asia and Africa.

Beyond its own borders, China is also funding or building more than 350 dams around the world [27]. Emerging as "contender to the power of western donors" [29], China is participating in at least $9.3 billion of hydropower projects across the African continent [35]. Companies like Sinohydro Corporation and Dongfang Electric Corporation financed by Chinese banks are investors behind the $2.2 billion Gibe III in Ethiopia (Africa's tallest dam), Egypt's $705 million Kajbar Dam, and Ghana's $729 million Bui Dam on the Black Volta River. In recent years, Chinese investors have been particularly active in neighboring countries along transboundary rivers such as the Mekong [27]. In Southeast Asia, the Three Gorges was completed in 2006, the Lao Nam Theun was completed in 2010, while over 40 GW of hydropower is now planned in the Mekong Basin and in East Malaysian Borneo a series of twelve mega-dams are under development. China is involved in building over 125 dams in Southeast Asia, representing 45% of all Chinese overseas dams [36]. "According to the Lao government's own figures, by the end of 2016 Chinese companies had signed up for US$6.7 billion worth of construction projects in the country" — some 30% of the total earmarked for Laos' Mekong basin, making Laos the third-largest market for China in the Association of Southeast Asian Nations (ASEAN) bloc [37].

The Intergovernmental Panel on Climate Change (IPCC) predicts that hydropower generation will double in China between 2008 and 2035, and triple in India and Africa over the same period [38]. At least 3700 major dams, each with a capacity of more than 1 MW, are either planned or under construction, primarily in countries with emerging economies. Experts find that "following a period of such relative stagnation during the past 20 years, the current boom in hydropower dam construction is truly unprecedented in both scale and extent [39, p. 162]."

Compounding Effects of the Contemporary Dam Resurgence

Seeing the mega-dam resurgence through this lens of major contemporary global dynamics has critical implications for understanding the impacts of the development trend itself. For instance, researchers across various fields are noticing that not only has the pace of hydropower growth been unprecedented, but the physical and cultural geography of where hydropower development is now happening is also unprecedented. And this geographic factor is causing major compounding effects on the impacts of our energy technology solutions.

First, the collective nature of these shifts has meant that much of this new energy infrastructure is being built in tropical and subtropical zones, where the global south's emerging economy demand is growing. These zones are also home to many of our most critically important tropical forests, important for their global carbon stores, important as sensitive, concentrated zones of ecological diversity, and critically important for their cultural significance as some of the last remaining areas of indigenous livelihood in the world [40]. Given the nature of where the dam resurgence is happening, there are enormous human, environmental, and cultural costs both locally and globally.

New evidence finds that the resurgence of the large-scale infrastructure projects through new land acquisitions in tropical and subtropical zones is directly and simultaneously inducing a resurgence of population displacement and dispossession [19, p. 1]. This is at a time when these very indigenous communities are more vulnerable than they have ever been to the implications of displacement due to rampant environmental degradation, climate change itself, and urban migration. In fact some studies suggest that besides energy security or regional cooperation one of the primary motivations for Chinese investment in dams in Southeast Asia outside of its borders is "to spare China's own rivers and avoid resettlement" since domestically the over-damming of Chinese rivers has already displaced over 23 million people and significantly affected water availability [36, p. 313].

This displacement is exacerbated by the fact that tropical rivers are critical to global food security. In tropical rivers of Africa, Asia, and South America, rainfall drives a periodic flood pulse fueling fish production and delivering nutrition to more than 150 million people worldwide [41]. The Mekong River Basin alone hosts one of the largest inland fisheries in the world, and the over 370 individual dam projects proposed for the basin will likely modulate this flood pulse, thereby threatening food security for already marginalized communities. The main tool for environmental governance and licensing in countries like Laos is local environmental impact assessment, which in most cases does not provide adequate technical information for, and thus has had minimal influence on, policy decisions.

China itself has been heavily criticized for lax environmental and social impact assessment standards at home. For instance, over 300 000 deaths have been reported due to dam failure in China, and it is believed that the devastating 2008 Sichuan earthquake was triggered by the province's Zipingpu dam [42]. Since 1949, 23 million Chinese citizens had been relocated for dam construction, and 6.5 million of those since 2000. Meanwhile, the Three Gate Gorges dam was decommissioned four short years after being built due to siltation [43], like many others, and data shows that dams in China underperform regard electricity output, due to increasing drought and water scarcity. Brazil is also heavily criticized for weak licensing regulation for large dams, and a poor impact assessment process, that was further simplified and weakened in 2012 [44]. Hydroelectric power is particularly damaging in the Amazon as larger reservoirs are needed to compensate for lowland topography. For this reason many Amazonian dams suffer from chronic siltation, which reduces electricity production, drastically affecting river ecology. Furthermore, seasonal flow Amazonian rivers means that many dams perform at only partial capacity. This lack of transnational basin-wide assessment often leads to disjointed project development with exacerbated impacts [45].

Second, hydropower's reputation as a low carbon energy solution has come under major scientific scrutiny in recent years. According to the latest science, reservoirs in different natural belts are responsible for different levels of emissions. In many rocky regions low on vegetation and population, such as in Iceland and other northern mountainous regions, the production of electricity from hydropower with temperate reservoirs is a net gain in terms of mitigating emissions from electricity production. In Asia, Africa and South America however reservoirs inundate tropical vegetation that decays, releasing masses of methane and soil carbon that can represent a net loss for mitigation.

While estimating emissions from hydroelectric generation is still an evolving field, there is broad consensus among the scientific community that methane production is a major concern for tropical freshwater reservoirs [46]–[50]. Major emission pathways for fresh water storage reservoirs include diffusion of dissolved gases at the air-water surface, methane emission from organic matter decomposition, and downstream dam emissions from degassing at turbine and spillway discharge points [47], [50]. Research now shows that among other variables, the geographic location of reservoirs has a significant impact on the organic matter storage, water temperature, and subsequent emissions through these mechanisms [50]. For instance, Fearnside highlights the example of the Curua-Una Dam in Brazil, where massive emissions from turbines and spillways mean annual green-house gas (GHG) emissions 3.6 times higher than would be emitted by the equivalent amount of diesel generated electricity, and these emissions levels are more than a decade after the dam's reservoir was inundated [51]. Fearnside and Pueyo conclude that "emissions from tropical hydropower in particular are often vastly underestimated and can exceed those of fossil fuel for decades [52, p. 384]."

Third, a major impact of the increasingly available deregulated private finance has led to a proliferation of projects that are largely managed outside the realm of international conditionality or regulatory oversight. In 2013 the World Bank reversed its two-decade old decision to turn its back on large hydropower investment, citing its improved impact assessment guidelines. The Word Commission on Dams (WCD) was established in 1998 by the World Bank and the World Conservation Union (IUCN) as an independent, multi-stakeholder body to review the effectiveness of large dams and to develop internationally acceptable criteria and guidelines for their planning and operation [53].

After WCD's establishment, the World Bank went from a low of just a few million dollars investment in dams in 1999 to about \$1.8 billion in 2014. However this still amounts to less than 2% of hydropower project investment today, given all of the other development finance avenues now filling the gap. Instead of acting as a primary investor, the World Bank has stated that it now "typically acts as a 'convener,' bringing other financiers to the table [54]." Research finds that this switch to private financing for projects with such massive externalities "derisks" megaprojects for the private sector. "Very often this means privatizing profits and outsourcing risks to the public [38]."

South-South investment trends noted above bode well for regional integration and set the stage for other forms of South-South cooperation, such as technical assistance and capacity development. However, the requisite institutional reform to regulate such development projects has lagged. Much southern development financing is not currently tied to human-rights progress, environmental impact standards, or democratic and participatory civil society stakeholder engagement. Nationally backed development banks such as the Brazilian Development Bank, China Development Bank, and the Development Bank of Southern Africa, or the Asian International Development Bank, the very banks now sopping up the hydropower investment gap we discussed earlier, "have abysmal records in terms of transparency and in terms of social and environmental safeguards [38]," and can be looked to for "alternative sources of finance that are perceived to be faster, come with fewer conditions and are more flexible" [29]. In many cases the companies conducting feasibility studies are also the same serving as financiers, builders, and regulators of projects, which "results in a blurring of lines between these role(s)" and raises issues of transparency [36, p. 322], [33].

International guidelines have always been far from perfect, as the World Bank case study showed, but the reduced financial involvement of international institutions allows project developers to ignore international concerns, with major implication since political attention often comes to communities most greatly affected by environmental risks only when larger national or international geopolitical forces come into play.

Defining Problems and Solutions

We argue that articulation of this confluence of global dynamics and their subsequent compounding effect on impacts helps to explain the fuller story of our large energy infrastructure resurgence, as well as our current dilemma. Local and global tensions are growing between civil communities and policy makers as decisions affecting resources, ecology, inhabitants, and industry are quickly being made with little public consultation or open analysis of alternatives, socio-ecological impacts, or land-use tradeoffs. Yet as shown, these are the communities most heavily affected by dam-related forest loss, displacement, and food insecurity.

Indeed, the activism space around hydro-development has become increasingly violent, with many high-profile murders and kidnappings being reported in the past ten years. Ironically, it seems in seeking to provide energy, climate, and social security, those are the very same securities jeopardized and in many cases eroded through such infrastructure projects [55]–[58]. Literature on the political economy of energy transitions suggests that rather than safeguarding marginalized communities from depravation, large-scale energy projects often serve to exacerbate existing social tensions and conflict, intensifying various manifestations of insecurity [55].

Furthermore, large-scale hydropower is often proposed as a tool for energy security, stimulating local economic development, or power export revenue through a low-emission renewable energy technology [44]. However recent research finds that national plans for greater energy security often overestimate the need for infrastructure and investment [59]. Rather, exploration of numerous contemporary dam conflicts, such as the Yacyreta Dam on the Parana River, along the border of Argentina and Paraguay, the Belo Monte dam of Brazil, the Tawang dams of Arunachal Pradesh, India, and the Mekong Dams of Laos show that the use of this win-win low-carbon development "narrative" can in fact disguise perverse incentives of state elites for construction, and perpetuate the imbalance of power dynamics among local and global actors [29], [30], [60]. The modern-day hydro-resource conflict can be framed as a reiteration of resource conflicts past and ongoing, proving waterscapes to be a new frontier in the local resource commodification and territorialization conflict [61].

Power dynamics and political economy play a key role in determining the winners and losers among different energy pathways, and in whose favor the trade-off between competing policy objectives weighs. In a state-led, investor-driven, donor-shaped policy context where state elites and international actors exercise imbalanced agency relative to constituents, the interests of the poor and the interests of the environment can be marginalized [62]. For this reason many civil society representatives and people from affected communities argue that the issue of land rights and access to rights must now more than ever be a core part of development planning, rather than sitting on the periphery. As such, the literature calls for increased focus on cultural politics — the institutions and relations of power among state and non-state actors that govern energy regimes and the outcomes they produce [63]–[65].

Returning to our initial discussion of the global resurgence of the large dam, if we see the trend toward large dams as part of this complex sphere, the issue of energy supply quickly becomes embedded in more imminent issues of rights and inclusion, necessitating critical reflection on our global, discursive definitions of "problems" and "solutions." Not addressing these key issues can lead to inaccurate, non-strategic policy-making and possibly lead to the assumption of false dichotomies between policy goals such as emissions reduction and poverty reduction [66]. High-modernist initiatives that orient themselves around problem solving without precedent of consensus on the very definition of the "problem" being solved run the risk of undermining their own objectives by predetermining the ways in which the "problem" can be conceptualized, discussed, and assessed. In ostensibly solving problems of energy demand and climate change, the hydropower resurgence may perpetuate even larger problems both at the local landscape and for global commons [62], [67].

In a World where Novel is Normal, Mundane is the New Radical

We contribute to the growing body of literature on sustainable energy transitions by placing the mega-energy infrastructure resurgence in the context of the confluence of global dynamics that have led to its development. From this perspective, we posit that truly sustainable energy futures will require more radical attention to the global dynamics and cultural politics that account for the power-play among actors, and more radical attention to our definitions of problems and their solutions, as opposed to a focus on technological innovations and financing them.

Critical issues such as power symmetry, land rights, representation, and participation have persisted for centuries, but rarely factor into energy planning policy in concrete central ways. The work ahead is thus to create formalized spaces for inclusion of a diversity of actors in the planning process, and for the exercise of rights to participate in that process. We suggest three ways in which local and international energy planning processes can be revised.

As Escobar's framework of cultural politics suggests, [65] first is the need to limit cultural dominance in the state's key institutions, especially those that create and implement development policy and local institutions that control access to rights. Addressing cultural dominance could encompass extensive legal reformation; the establishment of anti-corruption legislation that limits political interference and promotes merit-based employment and business contracting; and legislation that institutes regulatory bodies for investors and local industries that are independent, transparent, and accountable to the courts.

Second is a need to create spaces for, and to support diverse visions of, rights and what the exercise of rights means [65]. Even within one river basin, ideas of resource, subsistence, autonomy, identity, economy, and development can differ widely. Acknowledging and empowering non-dominant bio-cultural experiences of nature is a move towards peace with justice. Importantly, enclosure through restructuring resource use can have the same impact as enclosure through physical fencing [68]. So, seemingly inclusive solutions to environmental conflict that involve community management of forests, payment for ecological services, algorithmic river flow control, or other such initiatives should be approached thoughtfully and through truly participatory decision-making processes.

Third, while inclusivity is critical, the legitimate community of people who have rights to participate cannot be a foregone assumption in negotiation processes [63]. Creation of such a community will involve conscientious attention to the diverse and more nuanced expressions of agency (political, ecological, and cultural) that are important in identification of stakeholders for public participation and involvement. An organized civil society that acknowledges its own diversity will further support a broader representation in decision-making processes.

In their popular paper on the virtues of mundane science, Kammen and Dove themselves state that "the major obstacles to developing sound environmental practices are not principally technological, though expanding our research efforts in that area is critically important. Instead, the primary stumbling block is the lack of integrative approaches to complex systems and problems [69, p. 12]." Especially as large-scale energy infrastructure and technology is projected to dominate energy planning policy in emerging economies, we argue that there is no time to ignore the pressing and yet often overlooked issues of problem definition, inclusivity, and power dynamics. Addressing these seemingly mundane, yet fundamental, challenges may be the radical solution our global society needs.

Author Information

Rebekah Shirley is the Director of Research at Power for All and a Research Affiliate of the Renewable and Appropriate Energy Laboratory (RAEL) at the University of California, Berkeley. Email: rebekah.shirley@berkeley.edu.

Daniel Kammen is professor and Chair of the Energy and Resources Group, and a Professor of Public Policy, and Nuclear Engineering at the University of California, Berkeley. He directs the Renewable and Appropriate Energy Laboratory (real.berkeley.edu), and contributed to the 2007 Nobel Peace Prize awarded to the Intergovernmental Panel on Climate Change.

References

[1] G. Clark, *A Farewell to Alms: A Brief Economic History of the World*. Princeton, NJ: Princeton Univ. Press, 2009.

[2] J. Goldstone, *Why Europe? The Rise of the West in World History 1500-1850*, 1st ed. Boston, MA: McGraw-Hill, 2008.

[3] K. Pomeranz, *The Great Divergence: China, Europe, and the Making of the Modern World Economy.*, rev. ed. Princeton, NJ: Princeton Univ. Press, 2001.

[4] A. Korotayev, J.A. Goldstone, and J. Zinkina, "Phases of global demographic transition correlate with phases of the Great Divergence and Great Convergence," *Technol. Forecast. Soc. Change*, vol. 95, pp. 163–169, Jun. 2015.

[5] A. Korotayev, J. Zinkina, J. Bogevolnov, and A. Malkov, "Global unconditional convergence among larger economies after 1998?," *J Glob. Stud*, vol. 2, pp. 25–62, 2011.

[6] M. Spence, *The Next Convergence: The Future of Economic Growth in a Multispeed World*. New York, NY: Picador, 2012.

[7] United Nations Development Program, "The Rise of the South: Human Progress in a Diverse World," United Nations, New York, NY, 2013.

[8] A. Subramanian, *Eclipse: Living in the Shadow of China's Economic Dominance*. Washington, DC: Inst. of International Economics, 2011.

[9] P. Dicken, *Global Shift: Mapping the Changing Contours of the World Economy*, 7th ed. New York, NY: Guilford, 2015.

[10] M. de Paiva Abreu, M. Agarwal, S. Kadochnikov, M. Mikic, J. Whalley, and Y. Yongding, "The effect of the World Financial Crisis on developing countries: An initial assessment," The Center for International Governance Innovation, 2009.

[11] P. Wiseman, "Risks abound as developing nations lead the recovery," *Salt Lake Tribune*, Apr. 2, 2011.

[12] U.S. Energy Information Administration, *International Energy Outlook 2016*, 2016.

[13] B. Flyvbjerg, N. Bruzelius, and W. Rothengatter, *Megaprojects and Risk: Making Decisions in an Uncertain World*. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[14] International Energy Agency, "Southeast Asia Energy Outlook: World Energy Outlook Special Report." OECD, 2013.

[15] J. Rogelj *et al.*, "Paris Agreement climate proposals need a boost to keep warming well below 2 °C," *Nature*, vol. 534, no. 7609, pp. 631–639, Jun. 2016.

[16] S. Saran, "Paris climate talks: Developed countries must do more than reduce emissions," *Guardian*, Nov. 23, 2015.

[17] S. Chakravarty, A. Chikkatur, H. de Coninck, S. Pacala, R. Socolow, and M. Tavoni, "Sharing global CO2 emission reductions among one billion high emitters," *Proc. Natl. Acad. Sci.*, vol. 106, no. 29, pp. 11884–11888, Jul. 2009.

[18] S. Borenstein and K. Ritter, "Developing nations shift emissions stance in climate talks," Dec. 4, 2015.

[19] K. Otsuki, M. Read, and A. Zoomers, "Large scale investments in infrastructure: Competing policy regimes to control connections," in *Global Governance, Politics, Climate Justice and Agrarian Social Justice: Linkages and Challenges*, the Hague, Netherlands, vol. Colloquium Paper 32, 2016.

[20] M. van Ginneken, "A decade of sustainable hydropower development — What have we learned?," *Nepal Energy Forum*, Feb. 18, 2015.

[21] A. Ansar, B. Flyvbjerg, A. Budzier, and D. Lunn, "Should we build more large dams? The actual costs of hydropower megaproject development," *Energy Policy*, vol. 69, pp. 43–56, Jun. 2014.

[22] T. R. Berger, "The World Bank's Independent Review of India's Sardar Sarovar Projects," *Am. UJ Intl. Pol.*, vol. 9, p. 33, 1993.

[23] OECD. "Development Co-operation Report 2014: Mobilizing Resources for Sustainable Development," 2014.

[24] A. Gonzalez, "South-South investment: Development opportunities and policy agenda," *Trade Post*, Apr. 28, 2015.

[25] OECD, *Development Co-operation Report 2016*. Paris: Organisation for Economic Co-operation and Development, 2016.

[26] United Nations Conference on Trade and Development (UNCTAD), "World Investment Report 2014. Investing in the SDGs: An Action Plan," 2014.

[27] F. Urban, G. Siciliano, and J. Nordensvard, "China's dam-builders: Their role in transboundary river management in South-East Asia," *Int. J. Water Resour. Dev.*, pp. 1–24, Jun. 2017.

[28] M. Ken, *Emerging Capital Markets And Transition In Contemporary China*. World Scientific, 2017.

[29] P. Newell and J. Phillips, "Neoliberal energy transitions in the South: Kenyan experiences," *Geoforum*, vol. 74, no. Supplement C, pp. 39–48, Aug. 2016.

[30] L. Baker, P. Newell, and J. Phillips, "The Political Economy of Energy Transitions: The Case of South Africa," *New Polit. Econ.*, vol. 19, no. 6, pp. 791–818, Nov. 2014.

[31] M. Power, P. Newell, L. Baker, H. Bulkeley, J. Kirshner, and A. Smith, "The political economy of energy transitions in Mozambique and South Africa: The role of the Rising Powers," *Energy Res. Soc. Sci.*, vol. 17, no. Supplement C, pp. 10–19, Jul. 2016.

[32] F. Urban, *Low Carbon Transitions for Developing Countries*. Routledge, 2014.

[33] O. Hensengerth, "Chinese hydropower companies and environmental norms in countries of the global South: The involvement of Sinohydro in Ghana's Bui Dam," *Environ. Dev. Sustain.*, vol. 15, no. 2, pp. 285–300, Nov. 2012.

[34] C. M. Ashcraft and T. Mayer, *The Politics of Fresh Water: Access, Conflict and Identity*. Taylor & Francis, 2016.

[35] R, all Hackley, and L. van der Westhuizen, "Africa's friend China finances $9.3 billion of hydropower," **Bloomberg.com**, Sep. 9, 2011.

[36] F. Urban, J. Nordensvärd, D. Khatri, and Y. Wang, "An analysis of China's investment in the hydropower sector in the Greater Mekong Sub-Region," *Environ. Dev. Sustain.*, vol. 15, no. 2, pp. 301–324, Oct. 2012.

[37] "Powering Investment in Laos — Asean, Laos," *Thailand Business News*, Nov. 27, 2017.

[38] E. Gies, "Private funding brings a boom in hydropower, with high costs," *New York Times*, Nov. 19, 2014.

[39] C. Zarfl, A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner, "A global boom in hydropower dam construction," *Aquat. Sci.*, vol. 77, no. 1, pp. 161–170, Oct. 2014.

[40] J. Kitzes and R. Shirley, "Estimating biodiversity impacts without field surveys: A case study in northern Borneo," *Ambio*, Jul. 2015.

[41] J. L. Sabo *et al.*, "Designing river flows to improve food security futures in the Lower Mekong Basin," *Science*, vol. 358, no. 6368, p. 1053, Dec. 2017.

[42] S. Ge, M. Liu, N. Lu, J. W. Godt, and G. Luo, "Did the Zipingpu Reservoir trigger the 2008 Wenchuan earthquake?," *Geophys. Res. Lett.*, vol. 36, no. 20, Oct. 2009.

[43] J. Hays, "DAMS AND HYDRO POWER IN CHINA | Facts and Details," 2013.

[44] E. M. Latrubesse *et al.*, "Damming the rivers of the Amazon basin," *Nature*, vol. 546, no. 7658, pp. 363–369, Jun. 2017.

[45] C. D. Ritter, G. McCrate, R.H. Nilsson, P.M. Fearnside, U. Palme, and A. Antonelli, "Environmental impact assessment in Brazilian Amazonia: Challenges and prospects to assess biodiversity," *Biol. Conserv.*, vol. 206, pp. 161–168, Feb. 2017.

[46] J. A. Goldenfum, "Challenges and solutions for assessing the impact of freshwater reservoirs on natural GHG emissions," *Ecohydrol. Hydrobiol.*, vol. 12, no. 2, pp. 115–122, Jan. 2012.

[47] M. Demarty and J. Bastien, "GHG emissions from hydroelectric reservoirs in tropical and equatorial regions: Review of 20 years of CH4 emission measurements," *Energy Policy*, vol. 39, no. 7, pp. 4197–4206, Jul. 2011.

[48] C. Deshmukh *et al.*, "Physical controls on CH4 emissions from a newly flooded subtropical freshwater hydroelectric reservoir: Nam Theun 2," *Biogeosciences*, vol. 11, no. 15, pp. 4251–4269, Aug. 2014.

[49] V. Chanudet *et al.*, "Gross CO2 and CH4 emissions from the Nam Ngum and Nam Leuk sub-tropical reservoirs in Lao PDR," *Sci. Total Environ.*, vol. 409, no. 24, pp. 5382–5391, Nov. 2011.

[50] L. Yang, F. Lu, X. Zhou, X. Wang, X. Duan, and B. Sun, "Progress in the studies on the greenhouse gas emissions from reservoirs," *Acta Ecol. Sin.*, vol. 34, no. 4, pp. 204–212, Aug. 2014.

[51] P.M. Fearnside, "Do hydroelectric dams mitigate global warming? The case of Brazil's CuruÁ-una Dam," *Mitig. Adapt. Strateg. Glob. Change*, vol. 10, no. 4, pp. 675–691, Oct. 2005.

[52] P.M. Fearnside and S. Pueyo, "Greenhouse-gas emissions from tropical dams," *Nat. Clim. Change*, vol. 2, no. 6, pp. 382–384, Jun. 2012.

[53] World Commission on Dams, *Dams and development: A new framework for decision-making : The report of the World Commission on Dams*. London, U.K./Sterling, VA: Earthscan, 2000.

[54] "Hydropower," in *The World Bank Group A to Z 2016*. The World Bank, 2015, p. 85a–86.

[55] A. Simpson, "The environment: Energy security nexus: Critical analysis of an energy 'love triangle' in Southeast Asia," *Third World Q.*, vol. 28, no. 3, pp. 539–554, Jan. 2007.

[56] J. Barnett, *The Meaning of Environmental Security: Ecological Politics and Policy in the New Security Era*. New York, NY: Zed, 2001.

[57] M.J. Struebig *et al.*, "Anticipated climate and land-cover changes reveal refuge areas for Borneo's orangutans," *Glob. Change Biol.*, 2015.

[58] T. Doyle and M. Risely, *Crucible for Survival: Environmental Security and Justice in the Indian Ocean Region*. New Brunswick, NJ: Rutgers Univ. Press, 2008.

[59] R. Shirley and D. Kammen, "Energy planning and development in Malaysian Borneo: Assessing the benefits of distributed technologies versus large scale energy mega-projects," *Energy Strategy Rev.*, vol. 8, pp. 15–29, 2015.

[60] P.M. Fearnside, "Viewpoint–decision making on Amazon dams: Politics trumps uncertainty in the Madeira River sediments controversy," *Water Altern.*, vol. 6, no. 2, pp. 313–325, 2013.

[61] M. Nüsser, "Political ecology of large dams: a critical review," *Petermanns Geogr. Mitteilungen*, vol. 147, no. 1, pp. 20–27, 2003.

[62] J.C. Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale Univ. Press, 1999.

[63] A. Baviskar, "For a cultural politics of natural resources," *Econ. Polit. Wkly.*, pp. 5051–5055, 2003.

[64] A. Baviskar, *Waterscapes: The Cultural Politics of a Natural Resource*. New Delhi, India: Permanent Black, 2007.

[65] A. Escobar, "Difference and conflict in the struggle over natural resources: A political ecology framework," *Development*, vol. 49, no. 3, pp. 6–13, Sep. 2006.

[66] Y. Mulugetta and F. Urban, "Deliberating on low carbon development," *Energy Policy*, vol. 38, no. 12, pp. 7546–7549, Dec. 2010.

[67] J.L. King and K.L. Kraemer, "Models, facts, and the policy process: The political ecology of estimated truth," *Cent. Res. Inf. Technol. Organ.*, Jan. 1993.

[68] C. Corson, "Territorialization, enclosure and neoliberalism: Non-state influence in struggles over Madagascar's forests," *J. Peasant Stud.*, vol. 38, no. 4, pp. 703–726, Oct. 2011.

[69] D.M. Kammen and M.R. Dove, "The virtues of mundane science," *Environ. Sci. Policy Sustain. Dev.*, vol. 39, no. 6, pp. 10–41, Jul. 1997.

Krishna Sood

The Ultimate Black Box

The Thorny Issue of Programming Moral Standards in Machines

PRILL/ISTOCK

We are all familiar with the textbook moral dilemma: A car is driving on a road when a child darts in front of it. If the car swerves in one direction, it will hit a car in the oncoming lane. If it swerves in the other direction, it will hit a tree. If it continues forward, it will hit the child. The car is travelling too fast to brake. Each decision may result in death. While the scenario is an extreme case, drivers make life-and-death decisions on a daily basis. To an extent we have laws, rules, and etiquette to guide our driving. In other situations, a driver must rely on their internal moral compass. But what if a human driver is not making the decision, and an autonomous machine is? The scenarios present difficult choices and weighty decisions for humans, let alone for car manufacturers, developers, and engineers who must design machines with such decision-making capabilities.

If you believe the car manufacturers, tech titans, and the U.K. government, we can expect an exponential increase in driverless vehicles on our roads over the next five years. Driverless cars are already being tested on U.K. roads. This means that, inevitably, algorithms must be programmed to make consequential decisions in a manner that aligns with current laws and, where laws do not exist, our moral sensibilities. The authors of the *AI Now* report [1] state, "AI does not exist in a vacuum." It is deployed in the real world and has the potential to cause tangible and lasting impact. The driving scenario illustrates the conundrum developers face when launching software that must be equipped to make a moral judgment. Can they be expected to accurately pre-program moral-based decisions into autonomous machines? If so, whose sense of morality should prevail? There may be ethical dilemmas, lack of harmonized views, as well as bias that come into play. Programming moral standards into algorithms remains one of the thorniest [2] challenges for AI developers.

Moral standards are transient and far from absolute. Moral inclinations [3] may be conditioned on cultural norms. Cultural norms are neither

Digital Object Identifier 10.1109/MTS.2018.2826062
Date of publication: 31 May 2018

The concept of morality is thought to be a distinctly human construct, developed through social values and codes as well as an individual's experience.

universal nor immutable. Societal values and standards vary over time and geography. Acceptance of premarital sex, women in combat, homosexuality, and bans on slavery illustrate the seismic shift in values some societies have experienced in recent decades. If a software developer in the U.S. programs an autonomous vehicle to prioritize the safety of the passenger over animals on the road, could the outcome of the decision made by the same autonomous machine deployed in India be considered amoral where cows are considered holy and have right of way on roads? Should companies developing autonomous vehicles prioritize commercial gain over a utilitarian concept of safety? Research [4] indicates that buyers are less likely to purchase a car that prioritizes the safety of others over the occupants.

Moral decision-making is highly subjective and individualized. It is contextual, specific to the facts, and personalized depending on the experience, bias, and understanding of the facts of the person making the moral judgments. Individuals will react differently to reports that a Texan was not indicted for bludgeoning another man to death. If we learn that evidence suggests that the victim was raping a female, we may alter our position. When we learn that the man who committed the murder was the father, as was the case here, and the female was his five-year old daughter, the outcome may sit comfortably with our own moral compass. On the specific facts, the act of murder was defensible by State law [5] and the father called 911 in an attempt to save his daughter's rapist's life. The death was legally and, some may argue morally, justified.

Morality is a nebulous concept. It is best evidenced [6] by the reaction of an individual when faced with a choice requiring quick action. Such decisions are based on our own internal programming and made in split seconds. They are not always rational or logical — for example, a decision to jump in to save a drowning child at the risk of one's own safety or return to a burning building to save a family pet. Humans are opaque, flawed, and sometimes make bad judgments. With hindsight, we can analyze actions and interrogate after the fact. The truth, however, is that morality remains the ultimate black box. The same individual in the same situation except for one variant may make a different decision. Given the mutable and individualized nature of morality, is it feasible that a programmer can develop software with acceptably harmonized moral standards?

Morality is a uniquely human concept. Unlike other life forms, humans are considered to possess the singular capacity to judge their own actions and those of others. Helen Guldberg [7] writes that humans are not born with this ability but are conditioned to consciously make moral choices. While scientists have found that some animal species exhibit signs of a moral system, the concept of morality is thought to be a distinctly human construct, developed through social values and codes as well as an individual's experience.

The conundrum for software developers is that while AI technologies can equal and surpass human notions of computational intelligence (for example, Google's Alpha Go champion), morality remains a solely human domain. While machines, in particular, humanoids [8], may appear to possess reflective thoughts, the output is, at least with present technologies, a result of the data inputted, the rules used to train the algorithms, supervision, and iteration to achieve a desirable outcome.

This is more than an academic debate. Today, AI systems are deployed in a plethora of everyday decision-making scenarios, including autonomous machines, insurance premium settings, and recruitment practices. Companies deploying algorithms that make consequential decisions are hiring philosophers and social science majors to grapple with these quandaries. It is certain that, in the near future, more and more of us will be faced with the consequences of autonomous decision-making machines (trains, trucks, cars, buses) in our daily commute or school run. An autonomous machine must be equipped to make life-and-death decisions in a manner consistent with the law or, in the absence of laws, acceptable social norms. Adopting an engineer's problem-solving mindset, the solution would be to test the decision-making process among a sample group, determine the most frequently selected response, and train the algorithm accordingly. There are, however, ethical implications with imposing someone else's version of morality on another as well as the statistical probability that some of the time, the algorithm will produce an output that does not align with our expectations. The algorithms will be trained on the individual or collective bias of the sample group. The issue is that morality is deeply personal. In start-up companies deploying drones, for example, the founder's morals steer influences whether to deploy a machine for global good (weather, agriculture, or rescue missions) or to aid particular nation-states' defense, surveillance, or border control strategies. As consumers and as a society, are we prepared for a software developer, or a company, to be the guardian of our moral sensibility?

AI has the potential to greatly improve the human condition. While there may be arguments against the deployment of certain machines and uses of AI technologies, such as facial recognition in autonomous weapons, the underlying technology itself is neutral. Moral considerations must extend before deployment to the design and development of autonomous machines and AI techniques. Ethics should be a core subject in science, business, and engineering course curricula. Standards, both internal for companies developing technologies and external for industries where such technologies are deployed, will assist in setting governance frameworks and best practice. Data used to train the algorithm must be clean, correctly labeled, and reflective of a diverse and inclusive user base. Robust testing must be carried out prior to deployment and companies should have the processes to demonstrate, albeit internally, that the ethical implications of algorithmic decisions have been sufficiently contemplated and mitigation steps put in place.

The transitory and subjective nature of moral inclinations requires ongoing evaluation [9] and iteration of the algorithmic training to ensure that the output continues to resonate broadly with societal norms. Humans, however, are fallible, and morality is a human construct that is subject to change. Despite an engineer's best efforts to train and test an algorithm prior to release, there may be edge cases in which the outcome affronts our (individual and collective) moral principles. As a society, we are generally accepting of human error. We are less forgiving of technology. When presented with hard, ethical dilemmas, is it (morally) justified for humans to demand a higher standard of a mere machine?

In the near future, more and more of us will be faced with the consequences of autonomous decision-making machines.

Author Information

Krishna Sood is a senior lawyer and an expert advisor to the U.K.'s All Party Parliamentary Group on Artificial Intelligence. The views expressed in the article are the author's own.

References

[1] A. Campolo *et al.*, "AI Now 2017 Report," *AINOW*, 2017. [Online]. Available: https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf.

[2] B. Deng, "Machine ethics: The robot's dilemma," *Nature*, 2015. [Online]. Available: https://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881.

[3] J. Prinz, "Morality is a culturally conditioned response," *Philosophy Now*, 2011. [Online]. Available: https://philosophynow.org/issues/82/Morality_is_a_Culturally_Conditioned_Response.

[4] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science*, 2016. [Online]. Available: http://science.sciencemag.org/content/352/6293/1573?variant=fulltext&sso=1&sso_redirect_count=1&oauth-code=a7f278ef-c324-41a1-9c02-b89ad24f091d#ref-22.

[5] R. Goldman, "No charges for Texas father who beat to death daughter's molester," *abcnews.go.com*, 2012. [Online]. Available: http://abcnews.go.com/US/charges-texas-father-beat-death-daughtersmolester/story?id=16612071.

[6] "Moral machine—human perspectives on machine ethics," Moral Machine, M.I.T., 2017. Available: http://moralmachine.mit.edu/.

[7] H. Guldberg, "Only humans have morality, not animals," *Psych. Today*, 2011. [Online]. Available: https://www.psychologytoday.com/us/blog/reclaiming-childhood/201106/only-humanshave-morality-not-animals.

[8] S. Nasir, "Video: Sophia the robot wants to start a family," *Khaleej Times*, Dubai, 2017. [Online]. Available: https://www.khaleejtimes.com/nation/dubai//video-sophia-the-robot-wants-to-start-a-family.

[9] M.C. Elish, "Response to UK House of Lords Call for Evidence," Data & Society Research Institute, Written evidence (AIC0221). [Online]. Available: http://data.parliament.uk/written-evidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/70517.html, accessed Mar. 2018.

Joseph Carvalko

Defending Against Opaque Algorithmic Meddling in Free Elections

The British philosopher, G. E. Moore (1873–1958) in his seminal work *Principia Ethica (1903),* wrote that when considering what's good or bad, we should begin by asking two questions: ought the thing under inquiry exist? And, as it concerns an action, how ought we act? Although "good" in and of itself may well be indefinable, asking these questions has a common sense ring, especially when deciding the myriad ethical questions that surround new technology [1]. For example, Western nations have universally said "no" to germline modifications of the human genome, fearing that such actions might irreversibly affect the human race.[1] In other instances, the genie escapes before any international consensus can develop to stop it from becoming a reality. In 1945, atomic bombs decimated two Japanese cities killing nearly a quarter-million people. Only after a long string of hydrogen bomb detonations, over a period of nearly 20 years, were nuclear weapons banned as between the two major nuclear powers, the United States and the Soviet Union. The ban expressed a "right action."

Unlike permanent changes to the human genome or the proliferation of nuclear weapons, most technologies do not pose existential threats, but nonetheless call for regulation.

STATIONARYTRAVELLER/ISTOCK

Social media may be in this category. By 2019, more than half the population of Western Europe and more than a third of the population of the Middle East and North Africa will be using social networks [4], [5].[1] This represents a giant step toward actualizing an aspiration of the UN's Universal Declaration of Human Rights: "*a world* in which human beings shall enjoy freedom of speech." But, as social media serves to transform free speech the world over, a pervasive infiltration of the information highway is underway by individuals and entities using bots and human agencies to invade our privacy and channel extremist, hateful speech in propaganda-like campaigns bent on undermining democratic institutions [6].[2] Time has come to consider steps that balance privacy and speech rights with the right to choose leaders without foreign interference [7].

Broadly speaking, social media has been in existence for thousands of years, e.g., letter writing. Although the 19th century telegraph permitted two-way communication, its use was largely confined to business. It wasn't until the 20th century that

[1] Article 13 of the Oviedo Convention bans all genetic modifications to the human germline. Also see, [2]–[3].

[2] Alleged Russian political meddling documented in 27 countries since 2004 [6].

Digital Object Identifier 10.1109/MTS.2018.2826061
Date of publication: 31 May 2018

radio and later television allowed a communal connection between broadcaster and audience. National and international regulation quickly followed these early wireless successes. Nearly simultaneous with radio, telephone networks established information flow, mainly between two parties. Again, regulation followed. But, not until the 21st century did social media platforms make it possible for billions of people to both broadcast and communicate between themselves. Yet, virtually no regulation exists. Should it, when it's become clear that recent elections and democratizing events in the Middle East, Europe and the U.S., suggest that social media and liberal democracy don't always point in the same direction?

Evgeny Morozov, in his 2011 critique of the Web's political ramifications asks, "What if the liberating potential of the Internet also contains the seeds of depoliticization and thus dedemocratization?" [8]. We need look no further than the Iranian revolution of 2009, energized largely by Twitter and Facebook to protest what many Iranians considered a flawed presidential election [9]. But, as demonstrators messaged via Twitter, the Iranian regime also used the Web, flush with data, to identify protesters, via photos and associated personal information. The regime then widely disseminated propaganda, which when combined with shootings, tear gassing and arrests, put the restive population into a state of paranoia, which resulted in tamping down the marches [10].

Unlike quelling incipient revolutions, the 2016 elections in the U.S. and the U.K. were seemingly tainted by covert operatives, who commandeered social media platforms for purposes of altering the political result in two of the world's oldest democracies. In March 2018, The Guardian reported that "The data analytics firm that worked with Donald Trump's election team and the winning Brexit campaign harvested millions of Facebook profiles of U.S. voters, in one of the tech giant's biggest ever data breaches, and used them to build a powerful software program to predict and influence choices at the ballot box" [11]. This development has led U.K.'s Prime Minister, Theresa May, to call for the Information Commissioner to investigate the circumstances of one of the most egregious invasions of personal privacy in memory, which in this case apparently influenced the outcome of an election.

Democracies need both a free Internet and free speech, and judging from the election tampering that has occurred recently throughout Europe and the United States, time has come to consider instituting standards and global policies. These should be backed up by industry enforcement mechanisms, using both humans and AI, to defend against opaque algorithmic invasions of privacy and fabrication of propaganda. Social media platforms might begin by promising a robust transparency and accountability, where recognized international watchdog agencies can identify incidents of electoral maladministration, and insure that the aggrieved have an opportunity to be heard and to enjoin activities when justified, i.e., to stop the infringement of the right to choose one's political destiny without meddling. Yes, social media is a fact of life, but ought we act now to enforce electoral norms, what G.E. Moore likely would have agreed is undeniably "good?"

Author Information

Joseph Carvalko, based in Milford, CT, is Adjunct Professor teaching Law, Science, and Technology at Quinnipiac University, School of Law, and a Legislative Coordinator for Amnesty International. Email: Carvalko@sbcglobal.net.

References

[1] J. Carvalko, "Self-absorption, Where will technology lead us," *IEEE Consumer Electronics Mag.*, vol. 5, no. 1, 2016.

[2] "CRISPR–Cas9: A European position on genome editing," *Nature*, vol. 541, no. 30, Jan. 5, 2017; doi:10.1038/541030c.

[3] National Academies of Sciences, Engineering, and Medicine, *Human Genome Editing: Science, Ethics, and Governance.* Washington, DC: National Academies, 2017; https://doi.org/10.17226/24623.

[4] "Social networking across Europe a patchwork of penetration rates. Social network usage is high in the Netherlands, Italy, while France, Germany lag the region," *emarketer.com*, Jun 9, 2016; https://www.emarketer.com/Article/Social-Networking-Across-Europe-Patchwork-of-Penetration-Rates/1014066, accessed 3/13/2018.

[5] "Internet user penetration in Middle East and Africa from 2015 to 2020," *statistica.com*, 2018; https://www.statista.com/statistics/325703/middle-east-africa-internet-user-penetration/, accessed 3/13/2018.

[6] O. Dorell, "Alleged Russian political meddling documented in 27 countries since 2004," *USA Today*, Sept. 7, 2017; https://www.usatoday.com/story/news/world/2017/09/07/alleged-russian-political-meddling-documented-27-countries-since-2004/619056001/.

[7] The Editorial Board, "Russian meddling and Europe's elections," *NY Times*, Dec. 19, 2016; https://www.nytimes.com/2016/12/19/opinion/russian-meddling-and-europes-elections.html, accessed Mar. 13, 2018.

[8] E. Morozov, "The Net delusion: The dark side of Internet," *Public Affairs*. Perseus, 2011.

[9] "Editorial: Iran's Twitter revolution," *Washington Times*, Jun. 16, 2009; https://www.washingtontimes.com/news/2009/jun/16/irans-twitter-revolution/, accessed Mar. 18, 2018.

[10] L. Siegel, "Twitter can't save you," *NY Times*, Feb. 4, 2011; http://www.nytimes.com/2011/02/06/books/review/Siegel-t.html?pagewanted=all, accessed Mar. 18, 2018.

[11] C. Cadwalladr and E. Graham-Harrison, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach — Whistleblower describes how firm linked to former Trump adviser Steve Bannon compiled user data to target American voters," *The Guardian*, Mar. 17; 2018; https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

The Case of the 2013 Presidential Elections in Cyprus

ALEXSL/ISTOCK

Social Media in Election Campaigns

D. Ktoridou, E. Epaminonda, and A. Charalambous

Social media (e.g. Facebook, Twitter, YouTube) are today a regular form of communication used widely by individuals and organizations. More recently, social media are also used by political parties to communicate with voters. As a result, political parties engage in a new kind of conversation with voters, transforming campaigning into something more dynamic compared to what it was in the past. This engagement however varies by society. In some cases political parties simply send messages to voters while in others communication is more active. The current study provides evidence on the extent to which political

Digital Object Identifier 10.1109/MTS.2018.2826063
Date of publication: 31 May 2018

parties and candidates adopted and used social media tools as part of their campaign in the 2013 presidential elections in Cyprus. Interviews with social media officers of five political parties were conducted with the aim of uncovering the frequency and type of social medial usage in the elections. Results reveal that social media were primarily used for one-way communication rather than being a means of discussion and interaction between politicians and voters. Most candidates used social media merely for dissemination of news, images, political messages, and upcoming events. However, all parties recognized that the role of social media in presidential elections could be enhanced to allow more interaction between candidates and voters.

Social Media Potential

A new type of web technology popularly referred to as social media have opened up possibilities for enhanced online human-to-human interaction. The use of social media is spreading fast, from organizing events to sharing information, and from forming groups to running entire political campaigns. Social media allow users not only to seek information, but also to interact with others through online expression such as posting political commentaries on blogs and social network sites and sharing multimedia commentary [8].

Political leaders and parties recently began to use social networking to achieve political objectives [13]. A wide range of factors is attributed to the increasing use of social media for online campaigns and online electioneering both by politicians and by citizens. Parisopoulos *et al.* [14] cluster these factors into two major sets, supply driven and demand driven. Supply driven factors are those factors that relate to the use of social media by politicians, and demand driven factors are those that relate to the use of social media by non-politicians for political reasons.

Candidates can adopt many Internet tools for the purposes of communicating with constituents and voters in order to collect donations, foster community, and organize events. The available tools include social networking sites (SNSs). These are web-based services that allow individuals 1) to construct public or semi-public profiles within a bounded system, 2) to articulate [27] a list of other users with whom they share a connection, and 3) to view and traverse their list of connections and those made by others within the system [4]. Facebook profiles also give candidates a platform to publicize their support for a number of existing political groups, causes, and other candidates. In addition, they can post notes to their supporters and respond to comments on their individual web pages or walls [10].

One of the most well-known examples of leveraging social media for a campaign is the 2008 American presidential election in which President Obama used social media to his advantage, posting on Facebook, tweeting, and creating YouTube videos that disseminated his message much faster than any traditional marketing medium. However, globally many politicians don't seem to utilize the potential of social media. Khaldarova *et al.* [7] for example, suggest that Finnish politicians do not yet appear to be making that breakthrough push in use of online social media. While some politicians use online social media tools helpful for interaction with citizens, most politicians use social media as a dissemination tool rather than as a way to engage with voters. In addition, it seems that social media is often used as a campaigning tool and then quickly abandoned after the elections.

The picture seems to be similar in other countries too. A study in [16] showed that the Internet is not being used to its maximum potential by Indian politicians. Many of the party sites were not interactive nor were they adequately updated. Also, not many politicians used their sites to disseminate news and photographs to potential voters and several sites remained unchanged before, during, and after the elections. Johannessen [9] has also suggested that Norwegian politicians are still uncertain about how to communicate using SNSs and what communication through SNSs should mean for the political process. Further, according to [14], Greek politicians have not acquired a deep understanding of Facebook's potential. Most candidates just upload promotional material or announcements rather than seeking meaningful discussion with their supporters.

> **Most politicians use social media for short-term dissemination rather than long-term engagement.**

More recently, in the U.S. 2016 elections, social media seemed to have played a very important role. According to Frank Speiser, for example (quoted in [28]), the recent election was the "first true social media election." Social media's role in the election has been bigger than ever before [29] and it seems to have carried tremendous influence on the electorate [30]. According to [27], 35% of people 18 to 29 years old said that social media was the most helpful source of information in the 2016 presidential election. The two main candidates' use of social media differed significantly, according to [32], though. Donald Trump's posts were often focused outward, attacking others rather than talking about himself or what he would bring to the office of President. Hillary Clinton's approach was more about leveraging the next generation, and she occasionally interjected sarcasm and humor. Also, Trump's inflection tended to be simple and emotional, and Clinton's approach was far more traditional, and included paid Facebook ads that took advantage of the large audience and precision targeting capabilities to reach donors and voters. Overall, as [31] concluded, not only was "[s]ocial media's influence in this presidential election stronger than it has ever been," but it "will shape campaigns for years to come."

> Social network sites are transforming the way politicians, government and citizens engage with one another.

This study provides evidence for the use of social media in the 2013 presidential elections in Cyprus. Cyprus is a changing society that is quite developed socioeconomically. GDP in 2014 was U.S. $29 670 when adjusted by purchasing power parity, a figure that is 167% of the world's average [26]. Cyprus has a high level of information communication technologies (ICT) use –(the percentage of households utilizing a computer, either desktop, portable or handheld, according the Statistical Service of the Republic of Cyprus is 70.3% [22]. Of these users, the 93% are frequent users of the Internet, and two thirds of Internet users are connected with social media networks such as Facebook, Twitter, LinkedIn, YouTube, etc. [20]. With this background, the aim of this research was to investigate whether social media were utilized in Cyprus for the 2013 presidential campaign and, if yes, to what extent and in what ways.

Related Work

Social media is a term often used with terms like social network sites, social Web, Web 2.0, and user generated content (USG) [7], and they are defined as online applications, platforms, and media, which aim to facilitate interactions, collaboration, and the sharing of content [12]. Weber [23] adds that social media is the online place where people with common interests can gather to share thoughts, comments, and opinions. Whereas in traditional media such as newspapers, radio, and television, communication is one-way, social media allows everyone to publish and to participate in multithread conversations online.

Social media comes in many forms [25]: blogs, micro blogs (Twitter), social networks (Facebook), media-sharing sites (YouTube), social bookmarking and voting sites (Digg, Reddit), review sites (Yelp), forums, and virtual worlds (Second Life).

According to [6], social media sites exist under the conceptual umbrella of Web 2.0. A Web 2.0 application is a technology that allows for user-collaboration as well as User Generated Content (USG), which focuses on individual participation and content creation on Web 2.0 applications. Web 2.0 technologies are a collection of social media by which people actively form, organize, edit, integrate, and rate Web content [13]. Web 2.0 includes social network sites such as Facebook, which allows users to create profiles and establish connections with friends and acquaintances on the Internet [17], [18]. Other formats such as microblogs and video-sharing are also included in Web 2.0. Microblogs such as Twitter allow users to post short messages that are published online in real time. Video-sharing sites such as YouTube enable users to share user-created video and interact with other users in an online community. These technologies support group interaction.

Social Network Sites

According to [23], social networks sites are places where people with a common interest or concern come together to meet people with similar interests, express themselves and vent. What makes SNSs unique is not that they allow individuals to meet strangers, but rather that they enable users to articulate and make visible their social networks. On many of the large SNSs, participants are not necessarily looking to meet new people but they are more interested in managing relationships by maintaining contacts with old friends who are already part of their extended social network [12]. Moreover, [9] noted that the most popular SNSs are those that focus on user-generated content, participation, openness, and network effects. Social networking is not mainly about technology but about covering people's needs for access to and sharing of information, collaboration, and the creation of identity. Johannessen [9] also suggests that SNSs should be treated more as a cultural than as a technological phenomenon. To gain the benefits of SNSs, owners of information need to open their data, think in terms of collaborative production of ideas and content, and to share ideas with others in order to create better information.

SNSs are now transforming the political background of how politicians, governments, and citizens will engage with one another. The game of politics no longer belongs wholly to the professionals. Politics is now in the hands of ordinary voters, those who know how to make the best use this new infrastructure. The medium that rose to the forefront of this dispersal of power and mobilization of active political engagement in the last few years is Facebook.

Facebook — Founded in February 2004, Facebook is a social utility that helps people communicate more efficiently with others. It is a social networking site that facilitates interaction among like-minded people by sharing information through the digital mapping of real-world social connections. Not only it is the most popular social networking site, Facebook is one of the world's most trafficked Web sites with 1.7 billion active monthly users [21]. Facebook users can express themselves politically in various ways, such as by making online donations, encouraging their friends to vote, and posting graphics or status updates expressing political attitudes and opinions [8]. Facebook is one of the modern communication channels that have been used by many politicians to spread their ideas, influence opinions, and get voters to vote for them in elections. Essentially, Facebook is doing the old jobs of the campaigns, only smarter and faster [3].

Twitter — The phenomenon of online social network microblogging has become widespread over the last few years. The most popular microblogging tool is Twitter, which is a real-time information network that connects users to the latest stories, ideas, opinions, and news about what they find interesting. Twitter focuses on small events happening in users' daily lives and work activities, thus enabling them to share updates with friends, family, and co-workers [1]. It was launched in 2006 and it now has approximately 695 million registers users [21]. At the heart of Twitter are small bursts of information called Tweets. Each Tweet is up to 140 characters long. Users can see photos, videos, and conversations directly in Tweets to get the whole story at a tweet, and all in one place.

Online Group Membership

The greatest benefit that Facebook has provided to political candidates thus far is the means to mobilize and organize thousands of supporters [10]. Membership in a group provides necessary motivation and incentive to be politically informed. Facebook introduced the "groups" application in September 2004 as one of its basic features. Feezell *et al.* [24] noted that the application allows users to share common interests with each other by providing a common space where users can meet others interested in a particular topic, spread information about that topic, and have public discussions relevant to that topic. With the "Message All Members" feature, group administrators can reach out to entire membership rolls at any time. They can send out messages to members regarding meetings or campaign events, creating an actual contact list of voluntary supporters, a powerful means to organizing large-scale political and social movements [19]. Members can also invite friends to join a group or forward messages from campaigns.

In many ways, the ability for civic participation through Facebook can be attributed to the consolidation of information gathering and transmittance, the wealth of information that is easily accessible, and the flexibility users have in when information is accessed. In addition, online groups allow members to express their opinions through posts and to engage on many levels with the group discussion and information sharing [24]. The group application was one of the earliest and still remains one of the most pivotal features contributing to the interactive nature of Facebook.

> **Facebook facilitates targeting of advertising and marketing through political segmentation.**

Social Media and Citizens

On the other hand, there are several demand-driven factors that explain the increasing use of Facebook by citizens for political reasons. Facebook enables citizens to engage in meaningful political dialogue with candidates. It also gives amateur activists an easy way to connect with other citizens around the globe and helps them push their collective concerns to the top of political agendas [2]. Facebook's lack of geographic boundaries makes it easy for like-minded individuals to form large-scale communities, while technological advancements make it possible for nontechnical people to take a larger role in organizing and running such communities [14]. Users can join political groups, download candidate applications, and share their political opinions through the many communication tools on the site. Furthermore, users can view their friends' activities and comment on friends' posts, thus engaging in active conversation about political issues. Also, Facebook offers to its users opportunities to develop civic engagement skills with little to no additional time costs while simultaneously having access to a potentially large enough public to develop civic skills. Finally, Facebook enables its users to express themselves politically in various ways such as by making online donations, encouraging their friends to vote, and posting graphics or status updates expressing political attitudes and opinions [8].

Political Marketing and Advertising

Political marketing is about political parties adopting marketing principles, concepts, techniques, and strategies to achieve their goals and objectives. According to [11], all political parties can apply the technologies or strategies of marketing, such as market research, market segmentation, market orientation, and relationship marketing in order to achieve goals and objectives. In the same way that businesses market their products and services, politicians market themselves in order to win elections. SNSs can provide detailed demographic information about its users, which maybe the basis of market segmentation.

Political advertising includes all means and technologies required to attract public opinion, and, eventually, votes [25]. Facebook helps ease targeting through its Social Ads. These ads allow campaigns to segment the Facebook community by selecting users based on age, gender, education, interests, relationship status, keywords, and political views.

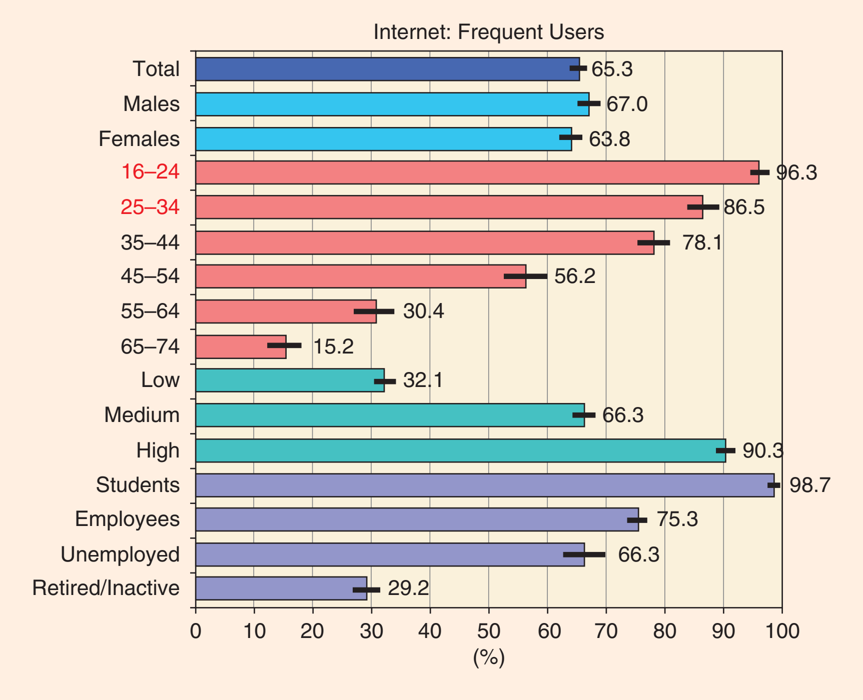


FIGURE 1. Internet users in Cyprus. *Source*: Statistical service of the Republic of Cyprus (2014) [22].

TABLE 1. Political parties' popularity on social media.

Party	Facebook (# of Likes)	Twitter (# of Followers)	YouTube (# of Views)	FLICKR	Instagram (# of Followers)
AKEL	6924	1603	21008	–	114
DISY	5462	2020	91944	√	–
EDEK	1602	712	6649	–	–
DIKO	1874	–	–	–	–
Green Party	233	400	7061	–	–
Citizens Alliance	2487	69	46670	–	–

Background Information

Internet Use in Cyprus

As Figure 1 shows, two in three people in Cyprus (65.3%) are frequent users of the Internet [22]. This percentage is very high at young ages (96.3% in the 16–24 age group and 86.5% in ages 25–34) and decreases with age, especially after 45, reaching 30.4% in the age group 55–64 and 15.2% in ages 65–74. High income individuals use the Internet more frequently (90.3% are frequent users compared to 32.1% in the low income group), whereas students use the Internet more frequently than workers (98.7% vs. 75.3%). A lower percentage of unemployed people (66.3%) are frequent users of the Internet, and a much lower percentage (29.2%) in the retired/inactive population are frequent users.

Most Popular Internet Activities

According to the same report, the most popular Internet activities among the Cyprus population are "Finding Information about Goods or Services," "Sending/Receiving Emails," "Participating in Social Networks," and "Reading/Downloading News/Newspapers/Magazines." As these results suggest, online social networking is a popular Internet activity, and political parties have a significant online presences.

Political Parties and Social Media

Political parties as entities also have accounts in social media. Facebook, Twitter, and YouTube seem to be used more extensively than Flickr and Instagram. Among political parties in Cyprus, Progressive Party of Working People (AKEL) has the highest number of likes on Facebook followed by Democratic Rally (DISY) and Citizens Alliance. On Twitter DISY has more followers than AKEL, and on YouTube DISY has the highest number of views followed by Citizens Alliance and AKEL (See Table 1).

Methods

In order to explore the uses of social media as online campaign tools in 2013 presidential elections in Cyprus, people in charge of social media communication in five political parties were interviewed. The five political parties were: Democratic Rally, Democratic Party (DIKO), The Movement for Social Democracy (EDEK), The Ecological and Environmental Movement of Cyprus and the Citizens' Alliance. DISY is a right-wing party that is currently the largest by vote in Cyprus. In the last parliamentary elections in 2011 it won 34.3% of the vote and holds 20 seats in the parliament. DIKO is a center-right party that is the third by vote, having won 15.8% of the vote in the last elections and holds 9 seats in the parliament. EDEK is the fourth party by vote. It won 9% of the vote and holds 5 seats. The Ecological and Environmental Movement or Cyprus (Green Party) took 2.2% of the vote and has one seat in parliament, while the Citizens Alliance did not participate in the 2011 parliamentary elections as it was founded in 2013. (Note: AKEL, the second political party by popular vote, was not included in the analysis because, despite efforts by the researchers, it was not possible to arrange an interview with a party representative.)

Interviews were arranged by phone in the period September–November 2014. Interviews were conducted in the participants' mother language — Greek — and the interview schedule was translated from English to Greek and then independently back-translated to check the wording. Interviews lasted from 30 to 60 minutes and were recorded and transcribed [5].

Questions asked related to which social media were used; reasons for using social media, their importance and benefits; effectiveness for targeting different age groups; marketing tactics to mobilize young people; and tracking young voters. Results are presented and discussed in the section that follows.

Findings and Discussion

Table 2 summarizes the responses of the parties' representatives to the questions posed at the interviews. As the table shows, all parties use Facebook, three use Twitter, two YouTube, and one Google ads. The fact that Facebook is the most used social media platform is probably not surprising given its high penetration and use.

A number of reasons are mentioned as to why political parties use social media. Some have to do with communicating more easily (e.g., "offer immediacy," "offer the ability to express political views and opinions online," "uploading promotional materials or political masseges," "interact with the others online"), reaching special age groups ("penetrate into age groups that we couldn't do through television"), and cost savings ("they are relatively inexpensive"). The benefits from the use of social media and their importance are linked to similar reasons, i.e., communicating more easily and cheaply, and reaching groups that are more difficult to reach with traditional media (four out of five party representatives agree with this last statement).

Persuading young people to vote for a candidate may be different than for the rest of the electorate.

Political parties employed a number of tactics to mobilize young people. These included using more bright colors, having advertisements that have their wording adapted for young people on YouTube, and promoting events or statements. The main argument was that young people use the Internet more and are convinced differently. According to one respondent: "We use different language for young people, more vivid, and more bright colors. We also designed some advertisements that played only on YouTube, not TV, because we felt this was more effective in reaching young people."

Only one party acknowledges that social media help them track voters; other parties claim that they do not. The "tracking" in the case of this party takes the form of following voters' comments and suggestions. As the representative of the party specifically stated, "Yes, [we track their views] since from there [social media] you can interact with the users and comment on activities or policies." One party representative expressed the view that tracking voters is illegal and unethical, apparently thinking about the practice of following as taking note of a voters' views.

Overall, social media seemed to have been used to a significant extent in the 2013 election in Cyprus. There does not seem to be significant differences between political parties in relation to the reasons for using and the benefits of social media. There was a difference in how representatives viewed "tracking voters," even though the kind of answers given may suggest that the meaning of the word "tracking" may have been understood differently by party representatives. In relation to young people, there is evidence that political parties treat them differently, acknowledging both young people's familiarity with social media and the fact that the tools needed to convince individuals from this group category may differ compared to the rest of the electorate.

Because of the exploratory nature of this study and the lack of comparative historical data in Cyprus, it is not possible to comment in relation to the changes that may have taken place in the last few years and the impact of the economic crisis on the use of social media in elections. The results of this study may be used as a benchmark for comparative studies in the future.

Further Exploration

Increasing Internet access in Cyprus has facilitated the use of social media for political purposes. In the 2013 presidential election campaign, social media were used quite extensively for dissemination of news, images, political messages, and upcoming events. Some parties and candidates had also begun to include newer technologies into their pages, such as Facebook application tools (e.g., group membership), Google, and Facebook ads. Even though social media, especially Facebook, Twitter and YouTube, were widely used, their usage involved primarily one-way communication instead of also having elements of two-way interaction with voters such as engaging in listening, responding to questions, or allowing interaction.

Future research in this area could explore the evolving use of social media for political purposes in Cyprus. Specifically, it might be studied how the economic situation after the 2013 elections may influence the use of social media in future elections. Differences between social media types and usage levels among political parties could also be analyzed. In addition, future studies can also analyze not only frequency of usage but also

TABLE 2. Summary of interview results.

	DISY	DIKO	EDEK	Green	Alliance
Which social media were used for promotion strategy	Facebook Twitter Google ads	Facebook Twitter YouTube	Facebook	Facebook Twitter YouTube	Facebook Ads Apps
Reasons for using social media	– penetrate into age groups that we couldn't do through television – they are relatively inexpensive – offer immediacy	– the ability to express political views and opinions online – interact with the others (online expression through posting)	– Dissemination of positions of EDEK – See what is written about EDEK	– To create online presence – uploading promotional materials or political massages	To directly communicate with people
Benefits for political campaigns through social media	you can locate and send your message in the age groups that you hardly can achieve through television	the direct interaction with young people	– no cost – divert new through social media – educate young people about the political reality	– Direct interaction with young people	Spread their campaigns using targeted tools and demographics
Why social media are important		– interact with anyone, anytime	– pass his/her positions in the wider public through social media	– Fast information spreading political messages about the political parties	
Are more efficient for targeting specific groups of voters than traditional media	Yes	No	Yes	Yes	Yes
Marketing tactic to mobilize young people 18–29 years old	– More bright colors – Advertisements on YouTube – Advertisement wording adapted for young people	Having a social media page	Promote events or statements		Hotspots, free Internet "MPOROUME" method
Do social media help track voters?	To some extent	No, it is illegal and unethical	No	No	Not track, grouping people

the content. Finally, the effect of social media on voters could also be the focus of analysis.

The use of social media across countries could further be compared, with the aim of identifying the factors that affect their level and type of use. Cultural differences between societies and the impact of the economic downturn maybe also studied. And, if the changes are in line with the trends observed in the recent American elections, social media will be an important aspect of electioneering, and related research findings would provide important information for academics and practitioners.

Author Information

Despo Ktoridou is Associate Professor and Head of the Management and MIS Department at the University of Nicosia in Cyprus.

Epaminondas Epaminonda is an Assistant Professor in the Management and MIS Department of the Business School of the University of Nicosia, Cyprus. He is the Associated Head of the Department and the Coordinator of the Doctoral Program of the School.

Andreas Charalambous is a University of Nicosia MBA Graduate.

References

[1] A. Aharony, "Twitter use by three political leaders: an exploratory analysis," *Online Information Rev.*, vol. 36, no.4, pp. 587–603, 2012.

[2] Anonymous, "Facebook and 'open source politics'," *Spreading the News*, 2007 [Online]. Available: http://spreadingthenews.wordpress.com/2007/10/05/facebook-and-open-source-politics/.

[3] N. Anstead and A. Chadwick, "Parties, election campaigning and the Internet: A comparative institutional approach," *Politics and International, Relations*, Working pap. 5, 2007.

[4] D. Boyd and N. Ellison, "Social network sites: Definition, history, and scholarship," *J. Computer-Mediated Communication*, vol. 13, pp. 210–230, 2008.

[5] D. Cohen and B. Crabtree, "Qualitative research guidelines project," Robert Wood Johnson Foundation, 2008[Online]. Available: http://www.qualres.org/HomeStru-3628.html.

[6] A. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business Horizons*, vol. 53, no. 1, pp. 59–68, 2009.

[7] I. Khaldarova, S. Laaksonen, and J. Matikainen, "The use of social media in the Finnish Parliament Elections 2011," *Media and Communication Studies Research Reports*, 2012.

[8] M. Kushin and M. Yamamoto, "Did social media really matter? College students' use of online media and political decision making in the 2008 election," *Mass Communication and Society*, vol. 13, pp. 608–630, 2010.

[9] M. Johannessen, "Genres of participation in social networking systems: A study of the 2009 Norwegian parliamentary election," *Electronic Participation*, vol. 6229, pp. 104–114, 2010.

[10] A. Sanson, "Facebook and youth mobilization in the 2008 presidential election," *Gnovis J.*, vol. 8, no. 3, pp. 151–174, 2008.

[11] L. Osuaqwa, "Political marketing: Conceptualization, dimensions and research agenda," *Marketing Intelligence & Planning*, vol. 26, no. 7, pp. 793–810, 2008.

[12] A. Palmer and L. Lewis, "An experiential, social network-based approach to direct marketing," *Direct Marketing: An International Journal*, vol. 3, no. 3, pp. 162–176, 2009.

[13] A. Attia, N. Aziz, B. Friedman, and M. Elhusseiny, "The impact of social networking tools on political change in Egypt's Revolution 2.0," *Electronic Commerce Research and Applications*, vol. 10, pp. 369–374, 2011.

[14] K. Parisopoulos, E. Tambouris, and K. Tarabanis, "Facebook and Greek elections: New fad or real transformation?," *IEEE Technology and Society Mag.*, vol. 31, no. 3, pp. 58–64, 2012.

[15] P. Rutledge "How Obama won the social media battle in the 2012 presidential campaign," 2013, [Online]. Available: http://mprcenter.org/blog/2013/01/25/how-obama-won-the-social-media-battle-in-the-2012-presidential-campaign/.

[16] R. Gadekar, K. Thakur, and P. Hwa Ang, "Web sites for e-electioneering in Maharashtra and Gujarat, India," *Internet Res.*, vol. 21, no. 4, pp. 435–445, 2011.

[17] K. Smith, *Social Media and Political Campaigns*. Trace: Tennessee Research and Creative Exchange, 2011.

[18] N. Smith, R. Wollan, and C. Zhou, *The Social Media Management Handbook: Everything You Need to Know to Get Social Media Working in Your Business*. New Jersey, U.S.A.: Wiley, 2011.

[19] N. Scola, "Despite negative press, Facebook is a powerful agent for social change," *Alternet*, 2008, [Online]. Available http://www.alternet.org/story/83196.

[20] A. Loucaides "5 things you ought to know about Internet usage in Cyprus," SocialWay, 2011, [Online]. Available: http://www.socialwayeservices.com/Our-blog/5-important-facts-about-internet-usage-in-Cyprus.

[21] *Statisticbrain*, Statistic Brain Research Institute, 2016 [Online]. Available: http://www.statisticbrain.com.

[22] Republic of Cypress, "Latest figures: ICT usage in households and by individuals," *Statistical Service of the Republic of Cyprus*, 2014 [Online]. Available: http://www.mof.gov.cy/mof/cystat/statistics.nsf/All/D0926894A2730949C2257D970035C525?OpenDocument&sub=3&sel=1&e=&print.

[23] L. Weber, *Marketing to the Social Web: How Digital Customer Communities Build Your Business*. New Jersey, U.S.A.: Wiley, 2007.

[24] J. Feezell, M. Conroy, and M. Guerrero, "Facebook is…fostering political engagement: A study of online social networking groups and offline participation," presented at APSA 2009 Toronto, *SSRN*, 2009 [Online]. Available: http://ssrn.com/abstract=1451456.

[25] D. Zarela, *The Social Media Marketing Book*. Sebastopol: O'Reilly, 2010.

[26] "Cyprus GDP per capita PPP," *Trading Economics*, 2015, [Online]. Available: http://www.tradingeconomics.com/cyprus/gdp-per-capita-ppp.

[27] J. Gottfried, M. Barthel, E. Shearer, and A. Mitchell, "The 2016 presidential campaign — A news event that's hard to miss," *Pew Charitable Research Center*, 2015. Available: http://www.journalism.org/2016/02/04/the-2016-presidential-campaign-a-news-event-thats-hard-to-miss/.

[28] J. Tsou, "Social media in the 2016 U.S. presidential campaign," *all about me*, 2016. Available: http://webkinzwate.blogspot.com.cy/2016/07/social-media-in-2016-us-presidential.html.

[29] S. Sanders, "Social Media's Increasing Role In The 2016 Presidential Election," *NPR.org*, Nov. 7, 2016. Available: http://www.npr.org/2016/11/07/500977344/social-media-s-role-increases-in-2016-presidential-election.

[30] M. Kapko, "How social media is shaping the 2016 presidential election," *CIO.com*, Sept. 29, 2016. Available: http://www.cio.com/article/3125120/social-networking/how-social-media-is-shaping-the-2016-presidential-election.html.

[31] M. Lang, "2016 Presidential Election Circus: Is social media the cause?," *San Francisco Chronicle*, Apr. 5, 2016. Available: http://www.govtech.com/social/2016-Presidential-Election-Circus-Is-Social-Media-the-Cause.html

[32] D. Kirkpatrick, "Is your social media strategy more Trump or Clinton?," *Industry Dive*, Oct. 20, 2016. Available: http://www.marketingdive.com/news/is-your-social-media-strategy-more-trump-or-clinton/428655/.

I Have Issues with Facebook

But I Will Keep Using It!

Ons Al-Shamaileh

Today, social media is one of the dominant channels for communicating and collaborating among individuals and organizations. People are using social media as an effective and inexpensive way to make friendships with new people and to keep in touch with the existing ones [1], [2].

Baruah [3] argued that social media has become an efficient and effective tool of communication in which individuals can share ideas and information, organizations can market their services and products, and customers can interact sharing their feedbacks and preferences. Hajli [4] introduced social commerce in his study in which the levels of interaction over social networks between consumers and businesses are beneficial.

Another motive to use social media is education. Zaidieh [5] showed that convenience, flexibility, and accessibility are the motives behind using social media in education; students can easily participate in discussions and share their experiences. Searching for jobs is another motive that has been presented by Black and Johnson [6] who showed that employers use social networking sites during the recruitment process.

With millions of hours spent on Facebook by the U.S. population, which is 18 times higher than the next biggest social network [7], Facebook is indeed an outstanding communication channel for millions of people. A study targeting 45 000 people aged 16–64 worldwide showed that Facebook has the largest share of social media users, where 82% of the study population has a Facebook account, and 42% of them are considered active users [8].

Shi *et al.* [9] investigated motives that influence users' intention to keep using Facebook; their results showed that keeping in touch with friends, entertainment and information seeking positively influenced users' satisfaction with Facebook. Additionally, Al-Menayes [10] presented entertainment, personal utility, information seeking, convenience and altruism as the main motives behind using social media.

Concerns

Nevertheless, despite Facebook's enormous popularity, concerns been raised previously in the literature.

Privacy

Privacy is one of the major users' concerns illustrated in the literature [11]. Stieger *et al.* [12] showed that privacy was one of the major reasons that led 48.3% of Facebook users to quit. Baumer *et al.* [13] mentioned that users are

Digital Object Identifier 10.1109/MTS.2018.2826078
Date of publication: 31 May 2018

KEVINJEON00/ISTOCK

not comfortable having their own life shown to the public. In addition, users stated that Facebook is not concerned about their privacy and did not trust Facebook securing their personal information [13]. Similarly, Fox and Moreland [14] showed that users did not like being unable to hide personal information from their existing network. In addition, users adjusted privacy settings because they were concerned about their privacy on Facebook [15].

Security

Security is another serious concern to Facebook users. In their study, Bilge *et al.* [16] proved how easy it was to attack a Facebook profile. Hackers steal identities and use them to send a friendship request to users, as a result, they gain access to the users personal information. For that reason, 69% of Facebook users are worried about the security of their personal information [17]. Another study investigated the reasons for quitting

Facebook in Japan and revealed that users left Facebook mainly because of security concerns, the requirements for declaring the real name, and the complexity of the Facebook interface [18]. Zheleva and Getoor [19] demonstrated how privacy attacks, with a mixture of public and private profiles can bypass the security settings in social networks. They also established evidence on how surprisingly easy it was to obtain private information from friendship links and group memberships on Facebook.

People find justifications for using Facebook despite their concerns over trust, privacy, and security.

Trust

Trust is yet another major concern about Facebook. Lankton and McKnight [20] distinguished between interpersonal trust beliefs, which include integrity, competence, and benevolence, and technology trust beliefs, which include reliability, functionality, and helpfulness. In his study, Deuker [11] argued that users did not provide Facebook with any information that needs protection, as they did not trust Facebook as a technology. Deuker [11] also revealed that users have doubts about Facebook privacy settings as well.

Annoying Content

Another dark side of Facebook is annoying content. Fox & Moreland [14] for instance, stated that users' reactions ranged from frustration to shock to revulsion at inappropriate content. Shelton and Skalski [21] also argued that Facebook includes negative content. Additionally, Butler [22] mentioned that risks increase with users' posting inappropriate comments on Facebook, where bullying can happen through text exchange and Facebook posts [23].

Usefulness, Usability, and Enjoyment

Previous research investigated the usefulness of Facebook as a concern. Among current Facebook users, 61% have voluntary taken a break from Facebook mainly because they believed that using Facebook is a waste of time and not useful as it contained too much gossip and drama, or they were too busy and not interested [24]. Similarly, it was shown that users' mood was negatively affected after using Facebook and users felt it was less useful and a waste of time [25]. Other studies reported usability issues with Facebook. Wang Y. *et al.* [26] provided examples concerning usability where feedback was not provided by Facebook when posting a video; thus a user did not know that she had accidentally posted a video until the next day. Another user could not delete a post on Facebook from his phone as he expected that the same functionality should operate on all platforms [26]. Similarly Morgan [27] mentioned that Facebook has major usability issues because of the targeted ads, sponsored status updates, and requests from third parties to get Facebook users to register for services. Li, Snow, and White [28] showed enjoyment as another issue and illustrated user views on Facebook with a user mentioning that Facebook is no more fun as it is all about chatting.

Control

A few studies addressed concerns related to control on Facebook. For example, Fox and Moreland [14] mentioned that users are not happy that Facebook is controlling their profiles. Facebook users were more concerned when the newsfeed option was released because they had less control over their information as the new option was an easy window for others to access their information [29].

So Why Do People Still Use Facebook?

With these investigations of Facebook concerns and issues in previous research, a question arises as to why people keep using Facebook if such negative perceptions and concerns exist?

Forty Facebook users were asked about the reasons behind their Facebook use despite concerns. The replies were analyzed using qualitative content analysis following the recommendations of [30]. The intention of using content analysis was to identify recurring concepts or beliefs of a salient issue within the answers, and to support any emergent issue from the replies [30].

Among the 40 respondents, 69% were female and 31% male. Thirty percent were between 18–24 years old, 40% aged 25–34, 20% ages 35–44, and 10% were over 44years old. The respondents' ethnicity was mainly Middle-Eastern 79%, followed by European 9%, Asian 5%, and 7% other/multi-racial.

In terms of educational level, 61% of the study population held a bachelor's degree followed by 19% Master's degree, 16% high school, and finally 4% were Ph.D. degree holders.

Respondents answers were grouped under relevant themes and reasons are presented in Table 1. The most recurrent reason for continuing to use Facebook related to perceiving Facebook as one of the most effective channel of communication.

Connection to Friends and Family

Respondents in this study confirmed their concerns about Facebook and provided justifications for using it despite the concerns. Most of the respondents stated that the main reason of using Facebook despite the concerns is to stay connected with their friends and family. A respondent stated "I do not like it (Facebook), but I am afraid that I will be disconnected from the people I know. Facebook is almost the only way to keep in touch with friends I have met from different countries, who I can no longer see in person." Another respondent was quite clear in stating the value of Facebook as an online, global channel of communication. She said

> "I like the idea of networking with my relatives and friends from all over the world. I (get so much) news from Facebook whether it is happy or sad; from a new born baby, engagement, wedding, etc. No matter what, Facebook is a good medium of communication."

Another respondent confirmed: "Facebook is the easiest and fastest channel through which I communicate with my extended family overseas." Another response that was aligned with the previous responses stated,

> "I still use it (Facebook) to stay in touch with friends and family, especially (since) many of them are living in other countries. And like it or not, people are not communicating with each other as they used to 10 years ago, but they take the time to post on Facebook."

One of the responses compared Facebook to other communication platforms:

> "The only constant communication way with friends and family who are faraway! Mobile numbers change so do addresses — Facebook accounts rarely change!"

These results align with the findings of Brandtzæg (2) who showed that people are using social media to make friendships with new people and to keep in touch with existing friends.

Education and Study Features

Another important reason for continuing to use Facebook relates to the features and services it offers, as one of the respondents confirmed: "Facebook is perhaps the most feasible way where I can connect with my colleagues instantaneously to discuss and exchange study-related matters." This agrees with Zaidieh (5) who illustrated the effectiveness of social media in education where students, educators, and administrators can effectively communicate.

Users cannot truly quit Facebook if they cannot permanently delete everything about themselves.

Commercial Uses

Commercial potential is another Facebook value mentioned by respondents: "I have a business page on Facebook. That is why I need it." Two other respondents concurred: "I have my own page and it is really beneficial," and "Facebook is the source of my income." These findings agree with Hajli (4) who discussed social commerce and how the levels of interaction between consumers in this type of e-commerce are beneficial over social networks.

Information Seeking

Considering Facebook as an information-seeking channel and a news platform was given as another reason for continuing to use the platform as stated by a respondent: "Facebook is a very effective channel to follow up

TABLE 1. Reasons for using facebook despite concerns.

No.	Reason
1	Effective and convenient channel for communication.
2	Utilitarian values of usage (e.g. study related purposes, commercial benefits).
3	Information Seeking (e.g., job search, job postings, news updates, etc.).
4	Potential loss of friends, connections, or acquaintances.
5	Enjoyment.
6	Virtual addiction.
7	Nonexistence of a better alternative.
8	Platform to express views and opinions.
9	Inability to entirely delete the Facebook account based on its terms and conditions of use.

on breaking news and events of interest." Another response that was similarly aligned:

> "Facebook is a news source for me. You find out about things happening in cities that won't even show on the TV news. Also, Facebook is a place to get information, inspirational stuff and recipes. Add to that groups which are very useful. You can connect with like-minded people and with similar background, communities, and nationalities."

Careers

Facebook value for job related purposes can be clearly seen in a respondent's comment: "The only reason keeps me using Facebook is the fact that many job opportunities are posted on it." Another respondent was also clear in asserting the benefits Facebook provides to users:

> "After using Facebook to get a scholarship, and getting a lot of money after advertising myself as an Arabic teacher for foreigners in Jordan, I do not think I will stop using it."

These results agree with Al-Menayes [10] who showed that information seeking, convenience, and altruism are among the main motives of using social media.

Lack of Alternatives

It is evident that users are aware of the benefits of Facebook, and some of them even perceive that no other social media network can compete with Facebook today. This was clear in one of the responses: "I will not quit Facebook, not until we have a better alternative."

Fear of Lost Connections and Virtual Addiction

Some respondents indicated that it will be difficult for them to leave Facebook. This was due to two reasons: First, due to concerns over potential losses:

> "I have built a huge social network of relatives and friends over the past few years using Facebook. It would very hard to stop using it now,"

a respondent acknowledged. Secondly, some respondents mentioned that they cannot leave Facebook because they feel they are addicted to it; as illustrated by two comments:

> "I am addicted to Facebook; it keeps me updated with what is going on around the world since I rarely watch TV"; and "Facebook is an addiction, simply an everyday thing."

This type of virtual addiction happens to people who become too reliant on their online identity as described by Modi and Gandhi [31].

Fun

Other responses highlighted the fun factor of Facebook, where one of the respondents focused particularly on the joy Facebook brings to her life: "I read my friends' posts on Facebook almost every hour. Many jokes around. So many funny videos." This finding agrees with Al-Menayes [10] who showed that entertainment is one of the motives for people to use social media.

Expressing Opinions

Considering Facebook as a platform to express views and opinions is another reason for people to keep using Facebook that was clearly seen in a respondent comment:

> "I consider Facebook as a platform to share my opinions and ideas, though sometimes I become worried since there is no[t] enough freedom where anything that you write can be legally used against you."

This finding agrees with Al-Saggaf [32] who showed that users expressed their feelings, shared their thoughts and opinions through Facebook.

Inability to Delete Facebook Profile Data

Finally, a respondent stressed out an interesting point where she mentioned that she cannot truly quit Facebook, even if she wanted so, since its terms and conditions of use do not allow the user to permanently delete everything about oneself. It is true that stopping the account of a person on Facebook will not delete the person's messages and photos; such information actually remains on Facebook servers after account deletion [33]. One of the respondents was aware of this fact and stated: "How can I really quit Facebook? My account will always be there. Facebook will not allow me to delete it even if I want to, so that is why I keep using it because it will always be there."

Further Research to Examine Cultural Factors Needed

A notable scarcity exists in finding studies which examine why people keep using Facebook despite their concerns. This research aims to help this problem.

Forty respondents provided reasons why they keep using Facebook despite their concerns. The majority of the reasons focused on the fact that Facebook is the most convenient means of communication that exists today, with great features and values that simply cannot be ignored. In addition, they mentioned that Facebook is one of the main channels for information seeking. Other respondents focused their reasoning on the fear of a potential loss of friends, not being able to entirely delete Facebook account, the joy Facebook provides to them, virtual addiction, the use of Facebook as a platform to express opinions, and the non-existence of a better alternative as perceived by the respondents.

In this study, the majority of respondents were female (70%), and respondents ethnicity is mainly Middle Eastern. More research is needed to examine the impact of culture on user concerns about Facebook with a more balanced gender distribution. In addition, further studies are needed to understand if users' concerns about Facebook could change over time.

Author Information

Ons Al-Shamaileh is an Assistant Professor in the College of Computer and Information Technology, American University in the Emirates, Dubai, U.A.E. Email: ons.shamaileh@aue.ae.

References

[1] A.N. Joinson, "Looking at, looking up or keeping up with people?: Motives and use of Facebook," in *Proc. 2008 SIGCHI Conf. Human Factors in Computing Systems*, 2008.

[2] P.B. Brandtzæg and J. Heim, "Why people use social networking sites," in *Proc. 2009 Int. Conf. Online Communities and Social Computing*. Springer, 2009.

[3] T.D. Baruah, "Effectiveness of social media as a tool of communication and its potential for technology enabled connections: A micro-level study," *Int. J. Scientific and Research Publications*, vol. 2, no. 5, pp. 1-10, 2012.

[4] M.N. Hajli, "A study of the impact of social media on consumers," *Int. J. Market Research*, vol. 56, no. 3, pp. 387-404, 2014.

[5] A.J.Y. Zaidieh, "The use of social networking in education: Challenges and opportunities," *World of Computer Science and Information Technology J. (WCSIT)*, vol. 2, no. 1, pp. 18-21, 2012.

[6] S.L. Black, and A.F. Johnson, "Employers' use of social networking sites in the selection process," *J. Social Media in Society*, vol. 1, no. 1, 2012.

[7] A. Lella, "Which social networks have the most engaged audience?," *comScore.com*, Apr. 2, 2015; https://www.comscore.com/Insights/Blog/Which-Social-Networks-Have-the-Most-Engaged-Audience.

[8] *GWI Social Summary: Global WebIndex's Quarterly Report on the Latest Trends in Social Networking, globalwebindex.net*, 2015; https://www.globalwebindex.net/hs-fs/hub/304927/file-2812772150-pdf/Reports/GWI_Social_Summary_Report_Q1_2015.pdf.

[9] N. Shi *et al.*, "The continuance of online social networks: How to keep people using Facebook?," in *Proc. 2010 43rd IEEE Hawaii Int. Conf. System Sciences (HICSS)*, 2010.

[10] J.J. Al-Menayes, "Motivations for using social media: An exploratory factor analysis," *Int. J. Psychological Studies*, vol. 7, no. 1 p. 43, 2015.

[11] A. Deuker, "Friend-to-friend privacy protection on social networking sites: A grounded theory study," in *Proc. AMCIS*, 2012.

[12] S. Stieger *et al.*, "Who commits virtual identity suicide? Differences in privacy concerns, internet addiction, and personality between Facebook users and quitters," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 9, pp. 629-634, 2013.

[13] E.P. Baumer *et al.*, "Limiting, leaving, and (re) lapsing: An exploration of facebook non-use practices and experiences, in *Proc. 2013 SIGCHI Conf. Human Factors in Computing Systems*. ACM, 2013.

[14] J. Fox and J.J. Moreland, "The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances," *Computers in Human Behavior*, vol. 45, pp. 168-176, 2015.

[15] E. Christofides, A. Muise, and S. Desmarais, "Information disclosure and control on Facebook: Are they two sides of the same coin or two different processes?," *CyberPsychology & Behavior*, vol. 12, no. 3, pp. 341-345, 2009.

[16] L. Bilge *et al.*, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proc. 18th Int. Conf. World Wide Web.* ACM, 2009.

[17] Rasmussen Reports, "69% of Facebook users concerned about security of personal information," *Rassussenreports.com*, 2010; http://www.rasmussenreports.com/public_content/lifestyle/general_lifestyle/may_2010/69_of_facebook_users_concerned_about_security_of_personal_information.

[18] A. Acar *et al.*, "Qualitative analysis of Facebook quitters in Japan," in *Proc. Eight Int. Conf. eLearning for Knowledge-based Society*, 2012.

[19] E. Zheleva and L. Getoor, "To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles," in *Proc. 18th Int. Conf. World Wide Web* (Madrid, Spain), 2009.

[20] N.K. Lankton and D.H. McKnight, "Do people trust Facebook as a technology or as a "person"? Distinguishing technology trust from interpersonal trust," in *Proc. 2008 AMCIS*. 2008, p. 375.

[21] A.K. Shelton and P. Skalski, "Blinded by the light: Illuminating the dark side of social network use through content analysis," *Computers in Human Behavior*, vol. 33, p. 339-348, 2014.

[22] K. Butler, "Tweeting your own horn," *District Administration*, vol. 46. no. 2, pp. 41-44, 2010.

[23] M.F. Catanzaro, "Indirect aggression, Bullying and female teen victimization: A literature review," *Pastoral Care in Education*, vol. 29, no. 2, pp. 83-101, 2011.

[24] L. Rainie, A. Smith, and M. Duggan, *Coming and Going on Facebook*. Pew Research Center's Internet and American Life Project, 2013.

[25] C. Sagioglou and T. Greitemeyer, "Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it," *Computers in Human Behavior*, vol. 35, pp. 359-363, 2014.

[26] Y. Wang *et al.*, "From Facebook regrets to Facebook privacy nudges," *Ohio St. Law J.*, vol. 74, p. 1307, 2013.

[27] E. Morgan, "Is Facebook suffering because of usability issues?," *usabilitygeek.com*, 2012; https://usabilitygeek.com/facebook-usability-issues/.

[28] J. Li, C. Snow, and C. White, "Teen culture, technology and literacy instruction: Urban adolescent students' perspectives," *Canadian J. Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, vol. 41, no. 3, 2015.

[29] C.M. Hoadley *et al.*, "Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry," *Electronic Commerce Research and Applications*, vol. 9, no. 1, pp. 50-60, 2010.

[30] M.B. Miles, A.M. Huberman, and J. Saldana, *Qualitative Data Analysis: A Methods Sourcebook*, 3rd ed. SAGE, 2014.

[31] Y Modi and I. Gandhi, "Internet sociology: Impact of Facebook addiction on the lifestyle and other recreational activities of the Indian youth," in *SHS Web of Conferences*. EDP Sciences, 2014.

[32] Y. Al-Saggaf, "Saudi females on Facebook: An ethnographic study," *Int. J. Emerging Technologies and Society*, vol. 9, no. 1, p. 1, 2011.

[33] Center, F.H. *How do I permanently delete my account?* 2017 [cited 2017 2/ January]; Available: https://www.facebook.com/help/224562897555674?helpref=related.

وعقيدة
تُخلع خالد المصلح أخلاق
سجاير المقطع
واحترام جديد موقع المصلح حجاب؟
المرأه العم ولا وتعتز مقابلة الزباله
الفتيات الله مدة مَتَاعًا يحجب عقل بها
غير نص قِل دينا او جمالنا نوع
بدون مو تعالى اللي ليست ،تُخلع تموتي ولادين إلى
تلبس برضا تخلع أن
حشيش هو عن جميل لا يُرمى وتعتز
هي الخيرر خمرة قلع الحجاب البنات وصباح دين
يعني مش ضيق لماذا السعوديات جميلات نظيف لأبناء
صورج فكر اي ،فكر مستقبل ما ،عقل لكن يرمى
او الولاد في يا ،جميل ف بلدنا
والخال تلبسي عند متحجبه
من هل يشتم بلا وبناء
الأنظار لبس بس رقي تتمسك انثى
،يُرمى حِجَاب شرعي
من حجاب ،يشتم
الحرية

Hijab in Twitter: Advocates and Critics

A Content Analysis of Hijab-Related Tweets

Mohsen Yoosefi Nejad, Mehdi Hosseinzadeh, and Maryam Mohammadi

One of the disputed topics in Islamic countries is the hijab, a veil that covers the head and chest, which is sometimes worn by some Muslim women. This study reports how Twitter users regard the hijab. As part of our investigation we collected hijab-related tweets from between August 28 and September 3, 2015, using NodeXL. We programmatically detected tweet languages using Language Detection API and categorized contents into eleven topics. Ninety percent of all tweets were in Arabic, seven percent in Persian and three percent in Urdu. Anti-extremism and hijab advocates were the most frequently

Digital Object Identifier 10.1109/MTS.2018.2826071
Date of publication: 31 May 2018

identified topics. Topics related to hijab and corresponding business goals, were infrequent, and rarely retweeted. We also calculated the impression to identify the most influential users, such as religious leaders, news agencies, and journalists.

Utility of Twitter for Investigating Social Phenomena

Online social media, such as the micro-blogging site Twitter, have become a rich source of real-time data representative of online human behaviors [1]. Unless the Twitter users mark tweets as private, tweets are public, providing information from a broad range of people on a variety of topics [2]. For example, Twitter is used for sharing information about healthcare, food consumption, social events, crisis, political views, sports, and culture.

Larsson and Moe [3] utilized online tools and presented a rationale for data collection and analysis of Twitter users, during the 2010 Swedish national election. Veenstra *et al.* [4] investigated Twitter use in citizen journalism in the 2011 Wisconsin labor protests. In this study, 775 030 tweets revealed differences in mobile and non-mobile use. Mobile users, who may have been present at the protests, posted fewer URLs overall; however, when they did post URLs, they were more likely to link to traditional news sources and to provide additional hashtags for context. Over time, all link posting declined, as users became better able to convey first-hand information.

In yet another study, collection of tweets posted from all local health departments (LHDs) having a Twitter account identified tweets related to the subject of diabetes in 2012 [5]. Using content analysis, the researchers grouped the diabetes-related tweets into categories and monitored LHDs' programs and trends. Ceron [6] used Twitter to predict Italian 2011 elections, French national 2012 elections, and the subsequent French legislative election. While Internet users are not necessarily representative of the whole population of a country's citizens, the Ceron analysis shows a remarkable ability for social media to forecast electoral results. In 2014 specific samples from the Twitter followers of the Canadian Football League led to insights about what motivates and satisfies Twitter followers of particular professional sport teams [7]. Selim *et al.*, [8] coded and analyzed about 5000 tweets of users from Saudi Arabia and the United Kingdom, to explore identity motives on Twitter. Their findings suggest that Saudi users appeared to seek the socio-psychological value of "distinctiveness," whereas British users appeared to seek out "belonging." In research conducted by Abbar *et al.* [9], the investigators examined the potential of Twitter to provide insight into U.S. dietary choices, by linking the tweeted dining experiences of 210 000 users to their interests, demographics, and social networks. Relating the caloric values of the foods mentioned in the tweets to state-wide obesity rates resulted in a correlation across the (fifty) 50 continental United States including the District of Columbia. The authors built a model to predict county-wide obesity and diabetes statistics, based on a combination of demographic variables and food names mentioned on Twitter.

Subject of Current Research

A topic of controversy in predominately Islamic countries surrounds women and the wearing of the hijab. A hijab is a veil that covers the head and chest that is worn by certain Muslim women beyond the age of puberty, in the presence of adult males outside of their immediate family, as a form of modest attire. The hijab can further denote any head, face, or body covering worn by Muslim women that similarly conforms to a certain standard of modesty [10].

The purpose of this research is to investigate the manner in which Twitter is employed in the context of issues regarding the use or wearing of hijab. In particular we endeavored to determine:

1) which segments of the population most tweet about matters concerning the wearing of the hijab; 2) categories identifying the main topics of veil-related tweets; 3) the degree to which Twitter users react to hijab commercial advertisement; 4) which groups of Twitter users influence the release of veil information.

Methods

The study represented was cross-sectional and descriptive. We sampled messages using the network analysis tool NodeXL [11] to collect tweets related to the keyword "hijab," symbolized by the character "حجاب," which has identical meaning in three languages: Arabic, Persian, and Urdu. We used percent encoding, according to RFC3986 [12] to supply our search term in NodeXL, as it does not support unicode characters as input. This resulted in 10 592 tweets, from which irrelevant messages were removed for a population of 6046 tweets for one week from August 28 to September 3, 2015. Using Language Detection API [13], the language of all tweets was detected. Figure 1 illustrates different language engagements compared with populations of people speaking in those languages in the world. Populations were extracted from Wikipedia [14], [15] and presented in Table 1. Excluding Urdu tweets, Arabic and Persian tweets were coded into 11 different groups. Each group implies a unique topic. These topics are displayed in Table 2. Figure 2 illustrates frequencies of tweets in each topic.

We identified the most influential users by calculating impressions [16], meaning the number of Twitter followers who potentially see a user's tweet. When a user tweets, these tweets are added to the streams of users who follow them, so that the tweet may be assumed to be seen and read [2]. Impressions were computed per user by adding the results of two expressions: 1) The first was calculated by multiplying the number of a user followers by the number of tweets he/she made; and 2) the second was determined by the number of followers who followed the posts of a specific user and retweeted by others. Combining topic clustering and impression factor, we extracted the top ten influential users in each topic group. Using different sources, such as their account on Twitter and other social networks, such as Facebook, Instagram, and the social intelligence platform, Klear [17], we identified the user type and classified them into four categories: regular users, religious leaders, news agencies/journalists, and special purpose users. The latter refers to accounts created for special purpose, such as supporting or criticizing religion.

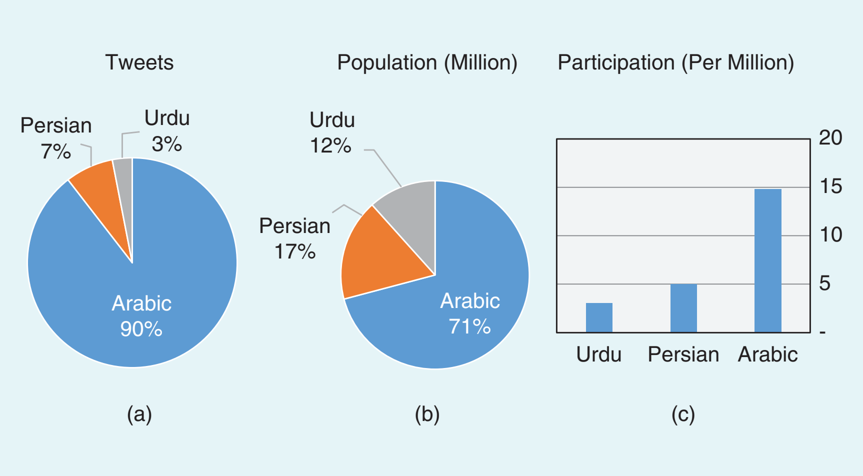


FIGURE 1. (a) Pie graph showing proportion of Arabic, Persian, and Urdu tweets. (b) Proportion of real world population of people speaking in different languages. (c) Bar graph showing participation ratio of Arab, Persian and Urdu society.

TABLE 1. User engagement categorized by language.

Language	Number of Tweets	Real World Population (Million)	Participation (Per Million)
Arabic	5412	365	15
Persian	449	90	5
Urdu	185	60	3

TABLE 2. Topics extracted from tweets.

Topic Code	Topic Description
T1	Hijab Advocates
T2	Covering the body is more important than covering the head
T3	Pictures of incomplete Hijab or Hijab-less women
T4	News
T5	Polling
T6	Hijab does not necessarily mean chastity
T7	Anti-extremism
T8	Hijab Critics
T9	Humors
T10	Advertisement
T11	Porn and misuse of offensive words for advertising

Results

There were 5412 Arabic and 449 Persian tweets regarding hijab, which means that Arabic tweets were 12 times as frequent as Persian tweets. Assuming the ratio of populations in the real world, the participation of Arab users is found to be three times as frequent as Persian users. Persian users tend to retweet news, while Arab users are more likely to advocate hijab or anti-extremism. Most of the news published by Persian users concerned a law voted for by the judiciary-cultural commission of the parliament of the Islamic Republic of Iran, which makes it illegal for a woman to remove her hijab while driving, and if the woman is found in violation of the law punishes her by a monetary fine (Figure 3(c)).

The most frequent tweet of Arabic advocates of hijab was a motto that says: "Liberty neither means dropping hijab, nor offending religion, nor eliminating morality, it is promotion of thought, respecting the mind, building a beautiful future, and a faith to believe and be proud of it." (Figure 3(a)). Most anti-extremists retweeted a strong objection to drawing hijab on the picture of a Syrian refugee mother with her child, who drowned at sea (Figure 3(b) and Figure 4).

We found that fifteen percent (15%) of Arabic tweets employed our search term used to determine the interaction with users and commercial websites. These tweets divided into two groups (i.e., T10 and T11). Fake hyperlink is a term we suggest for topics in common. Following the hyperlink provided in the tweet, a website will appear that is not related to the tweet. While the graph of the most frequent topics, T1 and T7, was regular, it seems completely irregular and noisy for T10 and T11 (Figure 5). This indicates tweets under topics T1 and T7 are retweeted or replied to by others, while fake hyperlinks rarely received engagement from users. In other words, marketers or robots reproduced fake hyperlinks, and then posted (not retweeted) these tweets, while useful tips are posted by a first person and then are retweeted by other users. The top ten influential hijab advocates were five regular users, two religious leaders, and three news agencies/journalists, while the top ten anti-extremists were five regular users, four news agencies/

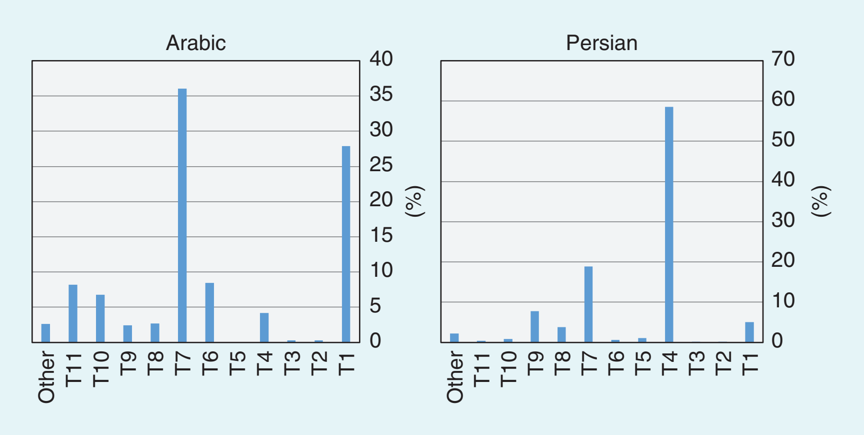


FIGURE 2. Frequencies of different topics in Arabic and Persian tweets.

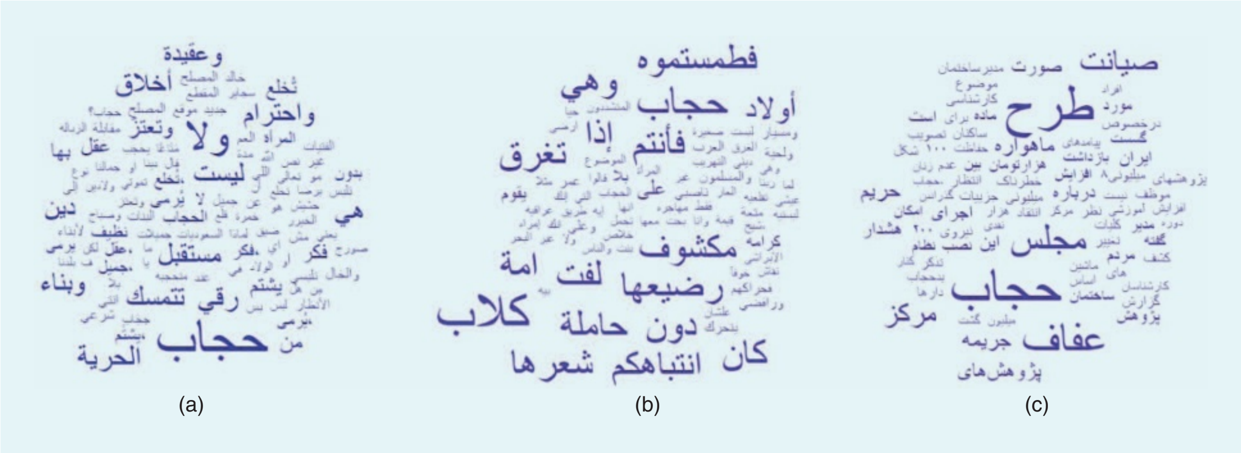
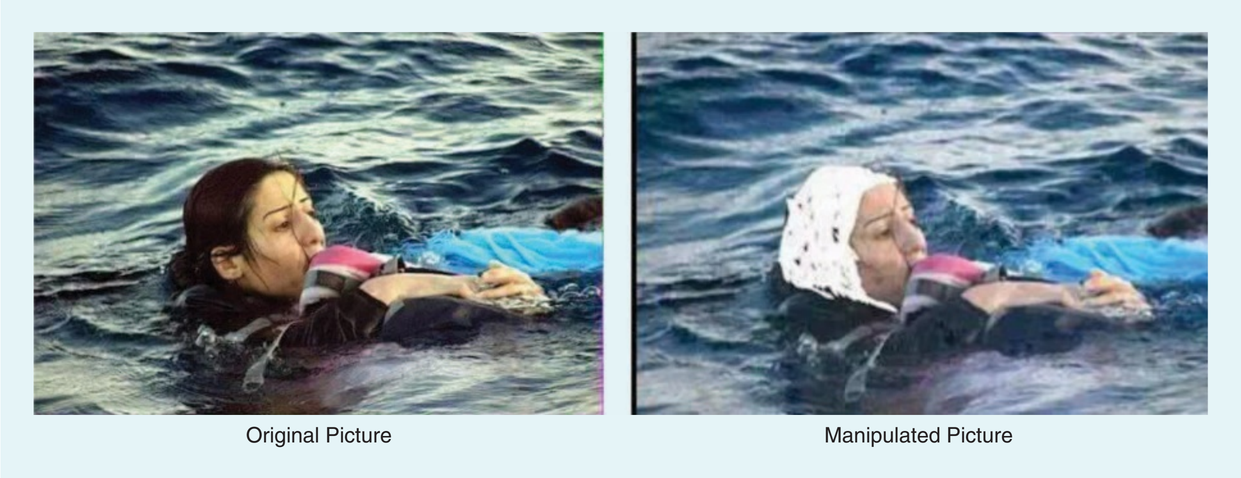


FIGURE 3. Word clouds of frequent topics. (a) Arabic T1 (Hijab Advocates). (b) Arabic T7 (Anti-extremism). (c) Persian T4 (News).


Original Picture

Manipulated Picture

FIGURE 4. Picture of the Syrian refugee mother with her child drowning in the sea.

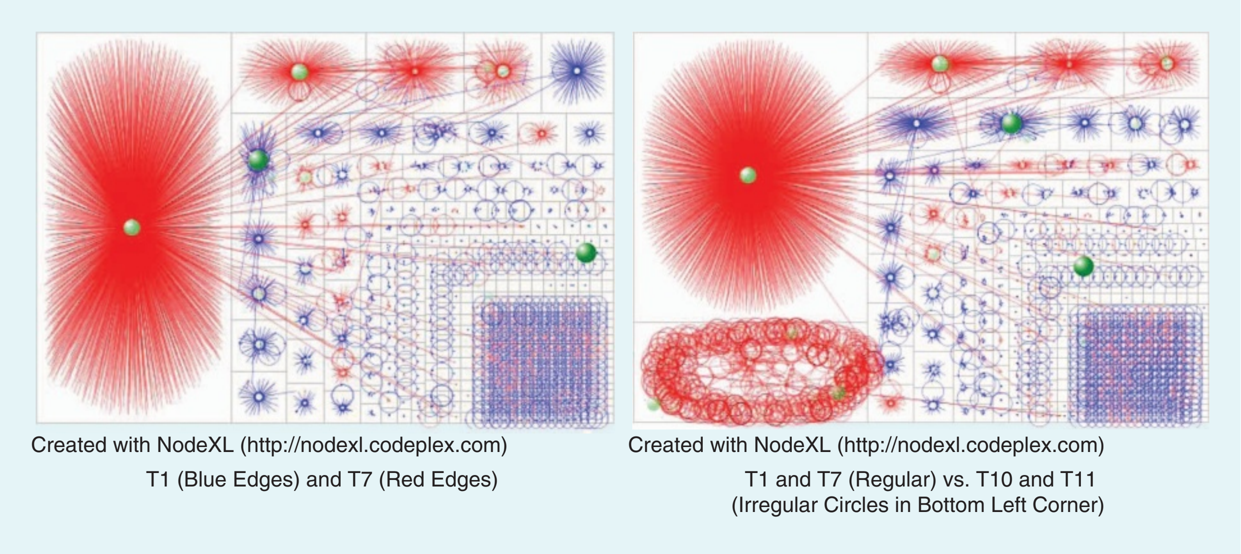

Created with NodeXL (http://nodexl.codeplex.com)
T1 (Blue Edges) and T7 (Red Edges)
Created with NodeXL (http://nodexl.codeplex.com)
T1 and T7 (Regular) vs. T10 and T11
(Irregular Circles in Bottom Left Corner)

FIGURE 5. NodeXL graph of popular topics in Arabic tweets versus unpopular frequent tweets.

journalists, and one special purpose user (see Table 3). According to the statistical results, it became clear that the top ten influential users were Arabic-speaking users.

Discussion and Analysis

Hijab is referred to by various names, the most common of which are: veil and headscarf. Most Muslims who wear the covering call it a hijab (حجاب). Islam introduced hijab to ensure decency and modesty where interactions between members of the opposite sex occur. While hijab is commonly associated with women, Muslim men may also wear a head covering as a show of modesty. Critics of the Muslim veiling tradition argue that women do not wear the veil by choice, and that they are often forced to cover their heads and bodies. In contrast, many Muslims argue that the veil symbolizes devotion and piety and that veiling is their own choice. To them it is a question of religious identity and self-expression. To this day, head coverings play a significant role in many religions. Christian and Jewish women in some traditions wear a headscarf as a religious, or cultural practice, showing a commitment to modesty or piety. Islam is known as a religion concerned with community cohesion and moral boundaries, and therefore hijab is a way of ensuring that the moral boundaries between unrelated men and women are respected. In this sense, the term hijab encompasses more than a scarf and more than a dress code. It is a term that denotes modest dressing and modest behavior.

This paper endeavors to provide answers regarding whether Twitter users are tweeting about hijab, who those users are, and the way such users frame their tweets. Although we detected the language of all tweets using Language Detection API, the relationships between spoken language and written language were found to be complex. The translation of written words requires significantly different processes compared to the translation of spoken words. In fact, automated systems, such as Google Translator and other similar translators are unreliable translators of spoken languages. This was a significant consideration for choosing our particular collection of tweets for analysis regarding the word "hijab" from Arab and Iranian tweeters. Additionally the keyword used "حجاب" has the same meaning in three languages: Arabic, Persian, and Urdu.

Because the number of tweets in Urdu was considered too small for analysis, they were rejected. This resulted in 5861 tweets for one week, August 28 to September 3, 2015. Obeying Islamic rules, including hijab

TABLE 3. Top ten influencing users in group T1 (Hijab Advocates) and T7 (Anti-Extremism).

	T1		T7	
User Type	Frequency	Mean Impression	Frequency	Mean Impression
Regular Users	5	7 928 540	5	4 770 645
Religious Leaders	2	7 881 121	0	–
News Agencies and Journalists	3	2 495 764	4	2 028 463
Special Purpose	0	–	1	7 254 734

or Islamic dress code, is required in Iran. We believe that this is the reason for the news published by Persian users concerning the laws of the judiciary cultural commission of the parliament of Islamic Republic of Iran regarding the wearing of hijab while driving. The most frequent tweet of Arab users regarding hijab was hijab advocates' and Anti-extremists' tweets.

We call attention to the fact that our study is constrained by five obvious limitations. First, we provided only a brief snapshot of a period of one week. Second, we limited our analysis to Arabic and Persian Twitter conversations. Third, messages were interpreted within the context of Twitter's 140 character format. Fourth, a significant weakness regarding the word-level analysis, is that detecting the use of irony and sarcasm is objectively impossible, as a fuller appreciation of context and motivation, among other factors, would be required. Fifth, we used the keyword "حجاب" in the search to capture conversations about hijab. This could potentially contribute to selection bias and failure to identify messages without this keyword that may have content related to hijab.

Advancing Social Media Studies

Hijab has always been a sensitive issue. Our survey confirms the existence of Twitter-based conversations about hijab use among Arab and Persian users. Findings from Twitter suggest that the number of Arab users were more than the number of Persian users in the use of Twitter regarding hijab and this is not due to the large population of Arabic speaking people. The most popular topics regarding hijab are hijab advocates' and anti-extremists' tweets. A NodeXL graph implies that users rarely retweet advertisements and fake hyperlinks. Impression factors indicate that religious leaders, news agencies, and journalists are the best known influencing agents in this area.

We believe that this work can be further expanded to examine the outcome of tweeting or reading about tweets regarding the hijab. Additionally, it can be expanded, for example, to evaluate the effect of participating in Twitter conversations about other sensitive cultural and religious issues in diverse languages. Our hope is that our work helps advance social media studies to identify trends within groups of users, especially for better understanding hijab customs and practices, where further research might be needed.

Author Information

Mohsen Yoosefi Nejad is with the Department of Computer Engineering and Information Technology, Payame Noor University, Iran. Email: m_yoosefi@pnu.ac.ir.

Mehdi Hosseinzadeh is with the Iran University of Medical Sciences, Tehran, Iran, and the Department of Computer Science, University of Human Development, Sulaimaniyah, Iraq. Email: hosseinzadeh.m@iums.ac.ir.

Maryam Mohammadi is with the Department of Computer Engineering, Islamic Azad University, Science and Research branch, Tehran, Iran. Email: mohammadi.maryam@srbiau.ac.ir.

> **Most anti-extremists retweeted a strong objection to drawing hijab on the picture of a Syrian refugee mother with her child, who drowned at sea.**

References

[1] J. Mathiesen, L. Angheluta, and M. H. Jensen, "Statistics of co-occurring keywords on Twitter," arXiv preprint arXiv:1401.4140, 2014.
[2] R. Thackeray *et al.*, "Using Twitter for breast cancer prevention: An analysis of breast cancer awareness month," *BMC Cancer*, vol. 13, no. 1, p. 508, 2013.
[3] O. Larsson, and H. Moe, "Studying political microblogging: Twitter users in the 2010 Swedish election campaign," *New Media & Society*, vol. 14, no. 5, pp. 729-747, 2012.
[4] S. Veenstra *et al.*, "Time, place, technology: Twitter as an information source in the Wisconsin labor protests," *Computers in Human Behavior*, vol. 31, pp. 65-72, 2014.
[5] J. K. Harris *et al.*, "Local health department use of Twitter to disseminate diabetes information," *Preventing Chronic Disease*, 10, 2013.
[6] Ceron *et al.*, "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France," *New Media & Society*, vol. 16, no. 2, pp. 340-358, 2014.
[7] N. Gibbs, N. O'Reilly, and M. Brunette, "Professional team sport and Twitter: Gratifications sought and obtained by followers," *Int. J. Sport Communication*, vol. 7, no. 2, pp. 188-213, 2014.
[8] H.A. Selim, K.M. Long, and V.L. Vignoles, "Exploring Identity Motives in Twitter Usage in Saudi Arabia and the UK," *Annual Rev. Cybertherapy and Telemedicine 2014: Positive Change: Connecting the Virtual and the Real*, vol. 199, pp. 128-132, 2014.
[9] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption through twitter," in *Proc. 33rd Ann. ACM Conf. Human Factors in Computing Systems*, pp. 3197-3206, 2015.
[10] G. Cyril, *The New Encyclopedia of Islam*. Walnut Creek, CA: Altamira, 2001, pp.179-180.
[11] Social Media Research Foundation, "Tools and Data For Social Media Network Insights;"http://www.smrfoundation.org, retrieved Oct. 15, 2015.
[12] T. Berners-Lee, R. Fielding, and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax. RFC 3986," 2005; https://tools.ietf.org/html/rfc3986.
[13] *Language Detection API*; http://detectlanguage.com, accessed Oct. 2017.
[14] "Arab World," *Wikipedia*; https://en.wikipedia.org/wiki/Arab_world, retrieved Sept. 15, 2015.
[15] "Persian People," *Wikipedia*; https://en.wikipedia.org/wiki/Persian_people, retrieved Sept. 15, 2015.
[16] W.K. Lindenwann, "Setting minimum standards for measuring public relations effectiveness," *Public Relations Rev.*, vol. 23, no. 4, pp. 391-402, 1998.
[17] Klear, *klear.com*; http://klear.com, accessed Oct. 2017.

The Social Metaverse

Battle for Privacy

Ben Falchuk, Shoshana Loeb,
and Ralph Neff

JUNGIN219/ISTOCK

Recent advances in technology are rapidly changing the way we interact with the physical world around us. As a result, our digital footprint and digital breadcrumbs are tracked and can reveal not just our identity but also our location, age, shopping preferences, friends, favorite movies, and much more. In the worst case, such tracking may lead to hostile entities coming to know your highly sensitive information such as credit card numbers, social security identity numbers, mother's maiden name, medical history, bank account information, and so on. Social engineering [1] is one of several related ways that this data becomes jeopardized. Furthermore, Internet-connected cameras allow consumers, companies, and government agencies to record animate and inanimate objects in a specific geographic area. Such recordings may be stored in cloud-based storage farms, viewed by humans, or analyzed by machines for various purposes. The information can be gathered and interpreted in multiple ways, such as by surveillance cameras, and can include activity and location inference as well as aggregation and pattern detection.

By and large, we are surveilled and sensed in many aspects of life. This includes: at home (e.g., smart grid energy monitors, ISP/Wi-Fi), while commuting (e.g., EZ-Pass, Google Traffic/Maps, fitness devices), in public spaces (e.g., public safety cameras and sensors, storefront cameras, webcams, etc.), and at work (company firewalls, corporate email, and Internet usage monitoring). In many cases, we are not even aware that such recordings and analyses take place and, hence, our privacy may be in jeopardy in ways we do not anticipate.

Digital Object Identifier 10.1109/MTS.2018.2826060
Date of publication: 31 May 2018

Here we distinguish several types of privacy (derived from [2]), including:

- Privacy of personal info: Any information that reveals something about physical, medical, physiological, economic, cultural, or social status.
- Privacy of behavior: Any information about habits, activities, choices, etc.
- Privacy of communications: Any data and metadata relating to personal communications.

Note that sometimes we accept a loss of privacy in exchange for security (in the case of security surveillance) or in exchange for useful customization (e.g., personalized advertisements). We also, sometimes unwittingly, freely offer up much of our personal information. For example, our mobile GPS location and device characteristics may be shared ubiquitously, and our social media posts may have a surprising reach (e.g., 150 000+ "friends of friends" [3]). Nowadays, as virtual reality (VR) applications increase in popularity and fidelity they also threaten to erode our privacy in new ways ranging from knowing how we physically move around to the patterns of our neural activities [4].

In this article we focus on technology underpinnings that will help VR participants increase the degree of privacy while immersed in social VR, and builds on our past research in privacy and gaming analytics [14], [15]. Though coined quite some time ago, we use the term *metaverse* with the same semantics as in Wikipedia [16]: "a collective virtual shared space, created by the convergence of virtually enhanced physical reality and physically persistent virtual space, including the sum of all virtual worlds, augmented reality, and the Internet." We use the term *social metaverse* to describe the above sorts of virtual realities in which a central purpose is socialization and interaction with other avatars — including both players and non-player characters (NPC's). Examples of software systems considered social metaverses today include: Facebook Spaces, AltspaceVR, Sansar, High Fidelity, and many more. While the social metaverse may or may not include capabilities such as gamification, realistic physics, realistic 3D models, user-created content, or in-game economies, it is the complexity and nuance created by the presence of other avatars (human or not) that most motivates our work.

Let us define the term "avatar" (or "agent") as a visible character within the social metaverse, constrained by the rules of the metaverse. We'll also use the term "user" (or "player") to connote a human who operates one or more avatars. Notably, the social metaverse:

- Is implemented by an engine that provides the computational basis ("rules of the game") for all aspects of the world including physics, appearance, communication, synchronization, etc. The engine is in sole control of the consistency and durability of the metaverse.
- Hosts avatars who cannot hide from the engine itself (if the engine attempts to surveil or analyze avatar activities, it may do so) nor can they perform actions not offered via metaverse API's.
- Is sometimes editable in the sense that avatars can affect the virtual world (e.g., create or destroy objects).

What is also true about the social metaverse is that, just like in the real world, those avatars who most skillfully use the capabilities of the world in the best way possible may experience a competitive or social advantage over others. We do not consider this to be a nefarious "gaming of the system" but simply using it better. Avatars, for example, may leverage a metaverse application program interface (API) and perform their own sort of surveillance and there is no guarantee that their actions or intents are ethically sound. For example, in-metaverse stalking is a dubious — but often allowable — kind of interaction with these worlds.[1] The metaverse will surely be underpinned by data analytics (DA) software components and combined with big data analytics and machine learning in order to provide the developer with insights into how users employ their services [4].

The stage, in our opinion, is therefore set for a battle for privacy within the social metaverse. While it may seem at present that little is at stake, one should note that it is possible that a good deal of our future lives may play out within these metaverses, including performing productive, meaningful work, exchanging important ideas, and using valuable digital currencies.

Motivation and Current Landscape

We have described how the stage is presently set for a privacy battle. This section provides more detail and some examples to corroborate this view. We also survey some of the academic work in this realm.

In the virtual reality metaverses seen in Hollywood movies — e.g., "The Matrix" and "Ready Player One" — participants often experience a level of fidelity indistinguishable from the real world [5]. The antagonists in these movies (but sometimes also the protagonists) often have special powers gained from their uniqueness or some sneaky shortcut. While the hacking of metaverse software underpinnings is an interesting field of

[1] There are increasing examples of inappropriate user behavior negatively affecting service offerings. Toxic players may insult others or threaten to "throw" the game or deliberately ruin other gamers' experiences in creative ways. In one example, toxic player behavior in the game Overwatch has delayed new levels and triggered serious re-thinking of processes for user management. (See: [17].) Other major entities in this space are both recognizing and starting to address toxicity-related issues. These include: Facebook [18], Steam [19], Google [20], and Rockstar Games [21].

its own [6], we neither consider it further here, nor require the presence of malignant insiders in order to justify this work. Current problems with identity theft, harassment, and more, within Massively Multiplayer Online Role-Playing Games (MMORPG's) further justifies our research. Related work on the causes and nature of in-game harassment (known as "griefing" in the realm of video games) indicates that social dominance orientation (a personality trait characterized by preference of hierarchical groups) is a strong predictor of online sexual harassment [7] and that lower-skilled male players are more likely to harass female players [8].

In 2014, the hashtag **#gamergate** mobilized a vast gamer campaign of ultimately criminal harassment (including threats of violence and rape) targeted at several women in the gaming industry. Within the game Second Life — an open world social metaverse — abuse and harassment were significant enough to warrant harassment "primers" by Linden Labs. One such primer offered the following advice: "If someone (or something) is pushing you or physically assaulting you inworld, sit down! Sitting prevents most physical forces from affecting your avatar" [9].

Another type of clear and present threat is that of social engineering hacking, a form of trickery that relies on human (victim) interactions that create a sense of urgency, fear, or other emotions, that lead to the individual revealing (unwittingly or not) something of value [1]. Avatars controlled by nefarious human users can easily engage in deceptive and unethical practices such as impersonation, white lies, and manipulation. For example, through observation over time an individual could impersonate a player's friend to obtain secret or private information. Such players might be annoying and, in the worst case, could jeopardize both player privacy and the pleasure of interacting with the metaverse.

Finally, while the present rapid advances of machine learning (ML) in various sectors — such as art, humanities, advertising, and chatbots — is paying dividends, there is also a potential darker side. Software-driven avatars — armed with ever-growing training data sets — can employ machine learning to nudge human avatars in ways that would best serve their purposes. When combined with social engineering this becomes a threat to privacy. For example, using in-game observations and logging, an ML-backed agent could come to know what your tendencies are, what kind of personality you have (such as impulsive, introverted, etc.), and what kinds of social interactions form the best "nudges" to create particular outcomes [10]. Furthermore, it will eventually be nearly impossible to differentiate between exclusively software-driven (e.g., chatbots, gamebots) and human-driven avatars. Indeed, detecting gamebots using analytic techniques is an active research field [11], [12].

Privacy Mechanisms in the Metaverse

Before describing some of our approaches to privacy we note that in the social metaverse all avatars must "play by the rules." What does this mean? Both the metaverse and the avatars are software, but the latter cannot exist without the former and the actual implementation underpinnings of the metaverse are accessible to avatars only through controlled means. An example of an access method into the underpinnings of the metaverse might be an API that allows an avatar to ask the metaverse for a list of other avatars presently within 100 distance units. In response the metaverse might return a list of avatars along with descriptive metadata such as skillset, interests, hometown, etc. Suppose that a direct API for listing nearby avatars was not available. It may still be possible for avatars to build up a similar capability through more primitive capabilities. For example, one capability might invoke a "snapshot" feature (to capture a rasterized image of the current scene from the avatar's point of view), and then call yet another module that picks out and enumerates the avatars in snapshots. Table 1 provides a summary.

We envision a new layer of controls that help tighten privacy in a metaverse where all avatars are essentially empowered by the same capabilities and act within the "rules of the game." Even with this assumption, however, significantly unethical, bothersome, and threatening behaviors might nonetheless still emerge. Consider that another avatar may: a) watch or follow you incessantly, b) monitor you from a distance, or c) harass you with its presence or utterances. The next sections address our approaches to mitigate these types of undesirable interactions.

Mechanisms

In this section we describe the mechanisms we believe will be useful in the battle for privacy. We view these as fundamental examples of tactics that will help improve privacy, but we recognize that other examples exist. These mechanisms are implemented in software and can exploit the primitives offered by the metaverse which we assume will include primitives that help enact movement, inventory, observation, and analysis of the metaverse. The broad goal of these mechanisms is to help ensure privacy while not utterly destroying the benefits of being in the metaverse. For example, while players could avert all threats to privacy by simply not entering a particular metaverse, this solution is too extreme to meet our requirements. The mechanisms should generally not come at the expense of participation in the metaverse or at the expense of interactions with other agents, objects, virtual storefronts, etc. To these ends, we define two important notions: privacy plans and confusion:

- **Privacy Plan:** A particular set of steps, initiated by an avatar, that enacts changes in the social metaverse such that the avatar has less risk of privacy intrusion when the plan is enacted. A plan can be thought of as a sort of program, written over the allowable metaverse API, that is carried out over a period of time.
- **Confusion:** Creating a confusing effect in nearby agents can be an essential part of an avatar's privacy plan. Whether or not nearby agents are human or non-player characters, a confusion tactic is intended to reduce the fidelity of these agents' knowledge of the avatar's activity, current or future position, possessions, interests, beliefs, and so on.

Note that in large social metaverses — as in MMORPG's — there is already a level of "cognitive load" introduced by the mere presence of other characters and such loads can ultimately diminish enjoyment of the experience [13]. Our work focuses on the more tangible and aggressive forms of privacy intrusions such as harassment and observation. The remainder of this section outlines privacy plans we have designed to help maintain privacy within the metaverse. Note that these are logical plans, not tied to any particular metaverse platform or specific app.

Plan A — Confusion — Creating a Cloud of Clones

In time, a complex metaverse will provide users with compelling reasons to want to confuse other avatars in their observable region. While interacting with other avatars will remain a principle pleasure (and main raison d'etre) of any social metaverse, it is likely that at times the sheer annoyance caused by some avatars (e.g., malicious strangers or bots), the sheer number of observing avatars, and the possibility of harassment or stalking (when another avatar simply follows you everywhere, essentially recording your experience) will make confusionary tactics attractive. One scenario warranting scrutiny is as follows: You are in a part of a metaverse that resembles a shopping mall in which many virtual (and real) products can be purchased at a multitude of storefronts. While each store may record your transactions, you may desire to obscure your movements from store-to-store from other avatars who you do not know nor trust. Why? For the same reason that an individual would not like to be followed shoulder-to-shoulder in a real mall while buying personal items, groceries, and books. Shopping habits can be highly predictive of other personal behaviors. In the metaverse it will be even easier to be observed by an annoying or malicious agent. Others can steer their avatars near yours, they see the view of the world that you do, and by following along with your avatar, observing and recording your avatars' interactions with others, store visits, and all other interactions that are observable, a detailed set of data about your habits can (in theory) be created. In another simpler scenario, you may simply be "hanging out" in the metaverse nearby a home you have created for yourself consisting of a building, a yard, and a lake. Here you may simply like to remain free from observation by peers while you stroll between the parts of your property — a reasonable desire indeed.

> It is necessary to ensure privacy without destroying the benefits of being in a metaverse.

We refer to one of our privacy plan classes as the "cloud of clones" plan. This plan's purpose is to bathe the environment with confusion in order to obfuscate user location, activities, beliefs, desires, and/or intentions. In this plan the system creates one or more avatar "clones" which have the same or similar appearance to that of the user's avatar. The clones may move about

TABLE 1. Aspects and responsibilities in the metaverse.

Aspect	Metaverse fabric	Avatars/players
Observation	Can observe and/or log any and all events within the metaverse.	Cannot hide from observation by the fabric of the metaverse. Can hide or obfuscate activity from other avatars in various ways.
Control	Controls all aspects of the metaverse including the application program interfaces that enable access to fundamental metaverse services.	Cannot perform activities disallowed by the fabric of the metaverse. Can potentially create surprising capabilities by "programming" them from primitive metaverse capabilities.
Ethics	Can put rules in place that help ensure ethical avatar behavior but cannot prevent all such behavior without reducing overall quality of experience.	As in the real-world metaverse credibility and reputation is tied to behavior but avatars are still free to behave in an ethically dubious fashion.

autonomously so that observers get confused and may not be able to tell which avatar is under the control of the actual human user. When clones are initiated the user may specify which behaviors are preferred for which subset of clones using command semantics such as: a) "assign all clones a behavior that has high randomness and high interaction levels," b) "assign half of the clones the behavior named 'walk around a house' and the other half of the clones the behavior named 'walk in circles.'" Behaviors may have additional configurable characteristics (e.g., the circles to be traced out might be 5 meters or 10 meters in diameter and/or might be centered on a specific location or on a specified object) and require specification of metadata such as: number of clones to spawn, duration of plan, spatial configuration, and more. Typically, a clone might closely resemble the user's avatar, but in principle, variants on clone rendition might include those that vary visually (and randomly) from each other (e.g., all wearing different colored virtual shirts or hairstyles). Each clone implements its behavior by performing in the metaverse, after which the plan terminates.

Figure 1 illustrates this paradigm in simplified form. Figure 1 (top panel) shows a stylized view of the metaverse in which our hero (avatar B) is near her virtual home. Two other avatars are very near to B, but B would like increased privacy from them. In Figure 1 (middle panel) B chooses, configures and launches the "cloud of clones" privacy plan. In Figure 1 (bottom panel) the plan executes, during which time the real B avatar eludes detection from A and C.

From an in-metaverse observer point of view (say Avatar A or C in Figure 1) the sudden emergence of a group of nearly identical avatars to the user (B in the figure) will create confusion. Importantly, the group contains the user B whose intent is to carry on with her actions without harassment.

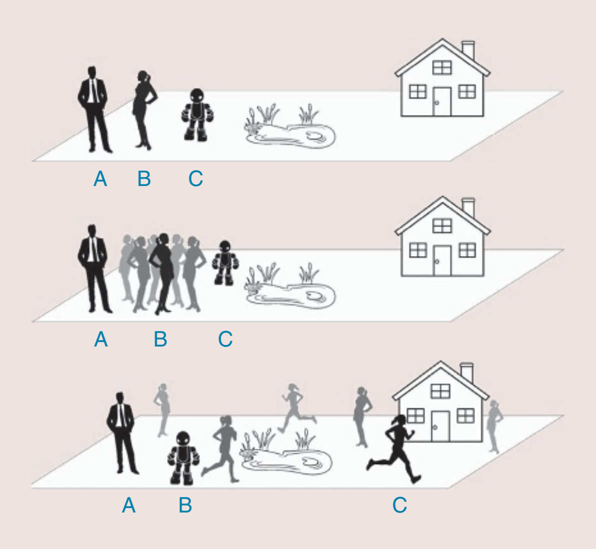


FIGURE 1. A privacy plan (involving clones) playing out over time.

It is desired that the sudden appearance and subsequent dispersal of these clones will cause any observers to lose track of the original "copy." During this time observers may be doing their best to track and analyze the behavior of B, but they would be forced to track all clones of B as well. To this end, the collective behavior of B and its clones is not as interesting when averaged out and it cannot be clear which behavior really typifies B's desires.

There are potential limitations to this approach which we continue to explore. For example, we presume that other nearby avatars cannot create a defense so as to disallow the creation of new clones, or that other avatars cannot (easily) use a method of locking in on a particular avatar (the original B) and tracking it programmatically. This latter possibility could undermine the sudden attempt at partial anonymity. To succeed over in-metaverse observers who may be identifying and tracking Avatar B in their viewports (the view of the metaverse seen from their avatar) the user might temporarily escape into an area in which observers cannot see him (e.g., a building or room) and execute the "clone" plan from there. So long as the observer does not gain visual access again before the clones are created the plan should be able to provide anonymity as desired. We note here that whether or not another agent could identify a clone by detecting pseudo-random behavior is an open issue. Finally, creating huge numbers of clones has effects on both performance and deployment that we do not pursue further here.

Plan B — "Private Copy"

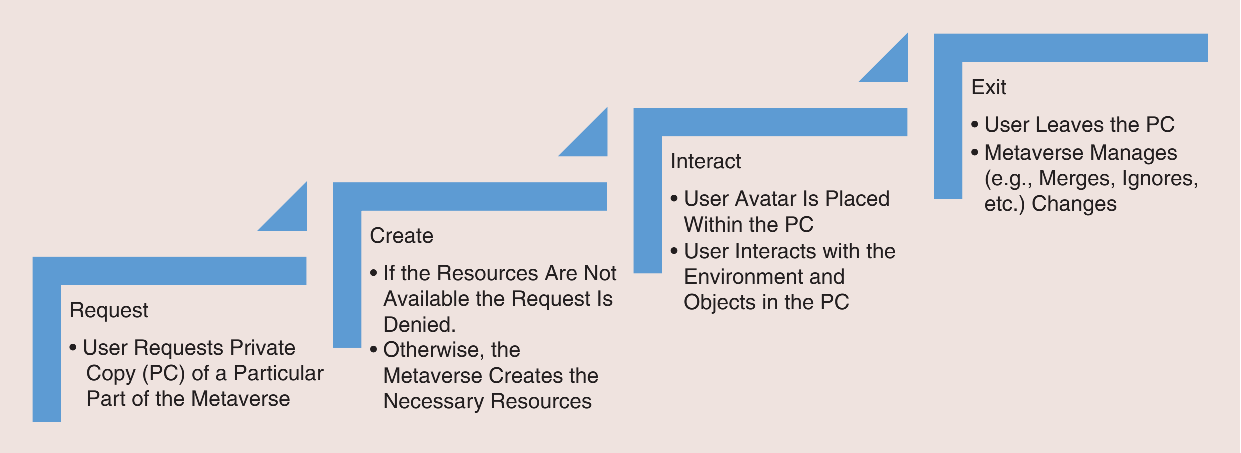
While the previous section proposes techniques to confuse surveilling avatars or bots, an alternative provides the user with a truly private space where surveillance cannot occur. In the current section, we discuss a class of privacy preserving plans which we call "Private Copy." In these plans, the user is able to request that a private copy of some part of the virtual world be created for the temporary exclusive use of that user. The corresponding portion of the metaverse in the main fabric continues to exist in parallel and other users and avatars may continue to use the main fabric portion unaffected by the actions of the user in the temporary Private Copy. For example, consider a user who desires a private virtual shopping experience. The user may request a Private Copy of a virtual store or even a portion of a virtual store (e.g., a particular department). For example, the store or department may sell personal items for which the user does not want to be observed shopping (e.g., virtual underwear, companionship services, etc.).

The metaverse will support an API from which the aforementioned user may create her space. A user interface supported by the metaverse will allow a user to request that the store or the department within the store be produced as a Private Copy. The Private Copy may either be created using resources on one of the metaverse provider's servers or in the user's client device. A "Private Copy" indicator should be visible to remind the user that the current experience is taking place in a private copy rather than in the full or otherwise more widely accessible virtual world. Once these steps are taken, the user may shop in the Private Copy of the store without worrying that other avatars or bots are observing. Back in the metaverse, from which the user originally triggered the Private Copy, the user's avatar might temporarily vanish, or a stand-in "clone" (see previous section) could mark the avatar's continued presence. The user interacts with the Private Copy for some amount of time, and then exits the Private Copy in order to return to the main fabric of the Virtual World. Figure 2 illustrates the main aspects.

Modifications to the virtual world itself may or may not be carried over from the user's interaction with the Private Copy. For example, the policy for a virtual world store may not allow avatars to make lasting changes to the environment in the store, and in this case the store environment always has the same appearance, according to the store provider's design. In this case, any environmental modification or interaction by the user within the Private Copy of the store would be necessarily discarded when the user exits the Private Copy. However, in some scenarios it may be useful to preserve modifications resulting from user interaction within the private copy. For example, suppose the user requests a Private Copy of a park within the virtual world, and then the user builds a gazebo in the center of the Private Copy of the park. The user enjoys the gazebo privately for some time, but then chooses to exit the Private Copy and return to the main fabric of the virtual world. At this point the system would assess the modifications the user made to the Private Copy, and would prompt the user to decide whether these changes should be discarded or preserved. If the user chooses to discard the changes then the Private Copy resources are freed and the user is returned to the (still gazebo-less) park in the main fabric of the virtual world. If the user chooses to preserve the changes then the system "merges" the changes into the main fabric, and thus the gazebo from the Private Copy may be added to the main fabric version of the park in the virtual world before the Private Copy resources are freed. In this case, the user and indeed any other avatar or bot will be able to see and to interact with the gazebo when present in the park. The gazebo would also be present in any future Private Copies spawned from the park by any user.

The nuances of merging changes from a user's Private Copy into the main fabric may depend on whether other avatars were present in the corresponding main fabric portion of the virtual world during the Private Copy session, and whether those avatars interacted with, modified, or observed that portion. In particular, it is possible for a modification made in a user's Private Copy to

> **Privacy mechanisms that do not preserve the continuity of the metaverse will not be acceptable to the users they intend to protect.**


Request

- User Requests Private Copy (PC) of a Particular Part of the Metaverse

Create

- If the Resources Are Not Available the Request Is Denied.
- Otherwise, the Metaverse Creates the Necessary Resources

Interact

- User Avatar Is Placed Within the PC
- User Interacts with the Environment and Objects in the PC

Exit

- User Leaves the PC
- Metaverse Manages (e.g., Merges, Ignores, etc.) Changes

FIGURE 2. The flow of steps for a Private Copy.

conflict with a modification which a different user made to the main fabric copy of the same portion of the virtual world. For example, a user enters a virtual kitchen room in which a knife is present on a counter. The user requests a Private Copy of the virtual kitchen, and while using the private copy the user picks up the knife and puts it into a drawer. Meanwhile in the main fabric version of the virtual kitchen, a different user picks up the same knife and adds it to his item inventory. When the first user exits the Private Copy of the virtual kitchen, suppose the first user requests his modifications to the virtual kitchen be preserved. In this case, the location of the knife must be resolved – is the knife in the drawer, or is it absent from the kitchen (removed to the item inventory of the second user?) In this case, it seems best to discard the first user's modification (as it was only witnessed by the first user in the Private Copy), and instead maintain the second user's modification. This is because the second user would be disturbed if the knife were to "disappear" from his item inventory, and also because additional avatars who observed the kitchen counter in the main fabric would have seen the knife removed by the first user. A change in the Private Copy will typically not have as many witnesses as the corresponding conflicting change in the main fabric. We have examined several ways in which conflicts can be resolved. For example (and not unlike the paradigms of software version control), the system may: a) adopt all changes from the copy into the main world, b) selectively merge changes from the copy into the main world, or c) preserve and merge changes from the copy only when they do not conflict with corresponding changes made in the main world.

A "Private Copy" plan thus gives the user absolute privacy for a limited time in a limited space of the user's choosing. While the user is immersed in the private copy, the system fabric guarantees that no other avatars or bots will observe the user's behavior. Of course, if the user chooses to merge private copy changes to the main fabric, it may be possible for an observer to later observe those changes and to deduce some part of the user's behavior. However, if the user discards changes made to the private copy, then no observable traces will be present in the main fabric, and the user's privacy will be fully maintained. In summary, we note that spinning up private copies of parts of the metaverse for small groups of avatars poses some IT-related issues such as scale and deployment, which are not detailed further here.

Toward a Framework of Privacy Plans

Once the system is capable of providing the user with a variety of privacy plans, it then makes sense to think of this set of plans as a privacy framework to be presented to the user in a controlled way. For example, the available privacy plans may be organized into a "Privacy Options Menu" that allows easy access to the various tools. Table 2 summarizes (in high level detail) several proposed privacy plan fundamentals that we have

TABLE 2. Various privacy plans.

Name	Summary
Clones	The fabric creates a "crowd" of new avatar clones, each identical in appearance to the user's avatar, in such a way as observers are confused and may lose track of the user's avatar.
Private Copy	Allows at least a portion of the VR world to be spawned as a "Private Copy" which the user exclusively inhabits and interacts with, unobserved by others.
Mannequin	The fabric replaces the user's avatar with a single clone of the user which exhibits believable behavior, while the user's true avatar is transported to another place. A mannequin is a type of clone that typically stands-in for a user's avatar whilst the user's attention is elsewhere.
Lockout	Allows a part of the VR world to be 'walled off' temporarily for private use; other avatars are temporarily locked out and prohibited from entering. For example, a room in a building may be subject to lockout so that a user has private use for some time. When the Lockout expires restrictions are lifted and other users are again allowed to enter the area and interact with this part of the VR world.
Disguise	Allows the avatar to stay in the local area but in a new (e.g. disguised) form. The disguised appearance may be randomly generated by the fabric, or the user may employ an avatar appearance editor to create one or more disguised forms for use with this plan. Observers do not easily notice the transformation and therefore become confused.
Teleport	The user's avatar is transported (e.g. instantly) to a new location in the virtual world. The destination may be selected by the user, e.g. from a list of destinations or by using a "map" interface.
Invisibility	The user's avatar takes an invisible form so that avatars and/or bots cannot detect the presence or actions of the user.

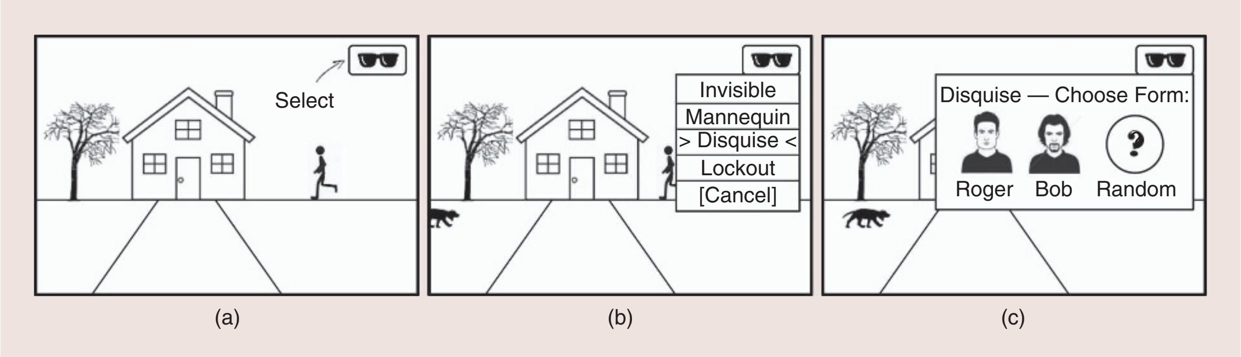


FIGURE 3. (a)–(c) A user selects one of multiple available privacy plans from a privacy options menu.

described in previous sections (or variants thereof). Though we lack the space to illustrate it, we believe that each of these fundamentals admits to a relatively simple algorithmic plan whose implementation would be useful to privacy-seeking avatars. Figure 3(a) shows the user's view of a local area of a virtual world. The privacy options menu is available through the user interface – in this example a sunglasses "button" appears in the upper right hand corner of the user's view. If the user selects this button, the menu of available privacy plans is displayed, as shown in Figure 3(b). The user then selects a plan, for example the 'Disguise' plan. In response, the system allows the user to choose the form of the disguise, as illustrated in Figure 3(c). Once the user has configured the disguise, then the user's avatar takes on the new disguised appearance, and in subsequent virtual world interactions the user's privacy will be preserved.

Once multiple privacy preserving tools are organized into a framework, it becomes possible to enable richer interaction with the available tools. For example, the framework may allow the user to select and execute multiple privacy plans together in useful ways. Here are some examples where a combination of privacy plans may provide increasing benefits over a single plan:

1) A user chooses to combine Disguise or Invisibility with Teleport such that the user's avatar vanishes from its original location and appears at a second location chosen by the user, but at the second location the avatar appears disguised or invisible (as previously chosen by the user). In this way, any avatars or bots observing the user at the original location will lose track of the user when teleport is engaged, and any users or bots at the second location will not "see" the user's true appearance and thus will not be able to identify the user's avatar at the second location.
2) A user chooses to combine the Clones plan with Invisibility such that when the additional cloned copies of the user's avatar appear, the user's true avatar simultaneously becomes invisible. Any avatars or bots observing the user before this combination of plans is executed will not only be confused by the appearance of multiple clones of the user, but will be guaranteed to lose track of the user's true avatar due to the invisibility effect.
3) A user chooses to combine Private Copy with Teleport so that the private copy of a part of the virtual world selected by the user is created, and the user is then teleported into that private copy. The portion of the virtual world selected for Private Copy may be quite distant from the user's starting location in the virtual world, for example the user may select this portion using a map-like interface, or perhaps select it from a list of identifiable locations from a menu. In this way, surveilling avatars or bots will merely see the user disappear for a time, and they will not see the user entering or approaching the location on which the private copy will be based.

Figure 4 illustrates a user interface for selection of multiple privacy plans in combination. In this example, the user begins in a park area of the virtual world, as shown in Figure 4(a). The user feels like gambling at a casino across town, but does not want to be seen at the casino. The user selects the privacy options menu as in the previous example, however now the menu allows the user to make multiple selections. In this case, the user selects both Teleport and Private Copy. Once the selection is made, the system allows the user to configure the selected privacy plans. As illustrated in Figure 4(b) the system may display a map interface and may ask the user to select the part of the virtual world on which the private copy is based. The user selects the casino from the map interface. With configuration complete, the system spawns a private copy of the casino and teleports the user into the private copy. The user's new view is illustrated in Figure 4(c). Note that some visual indication is given in the user's view to remind the user that he is no longer interacting with the full virtual world, but rather is within a private copy of the casino. The user enjoys gambling at the casino for a time,

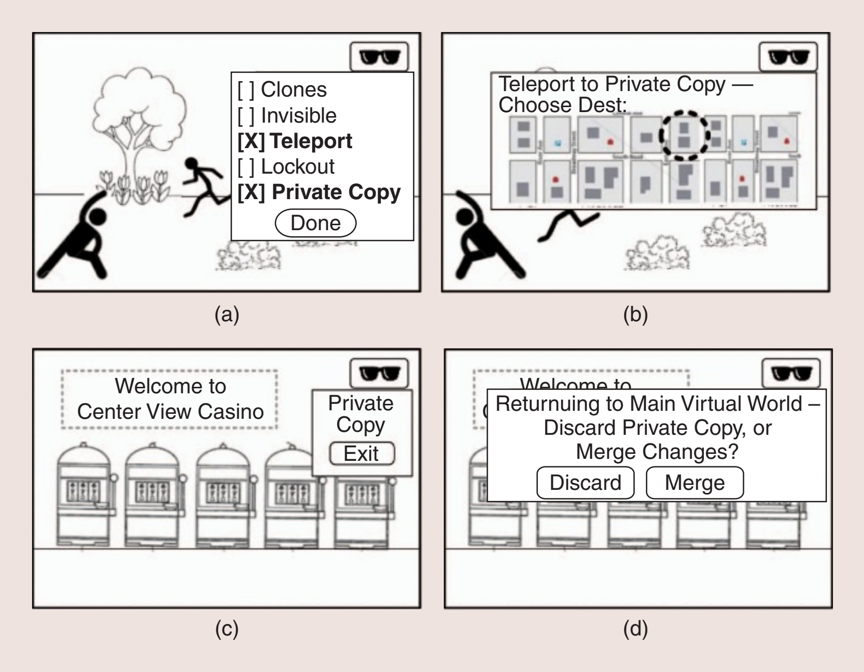


FIGURE 4. (a)–(d) User selects multiple privacy plans for use in combination.

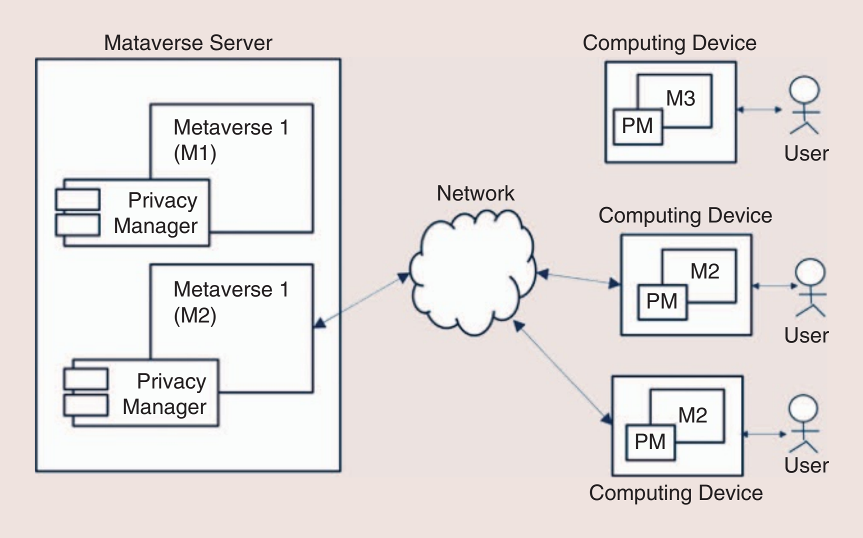


FIGURE 5. IT deployment in context.

and then chooses to exit back to the main virtual world. As illustrated in Figure 4(d), the system offers to discard or preserve changes that the user made to the private copy of the casino.

An additional benefit of having a framework of privacy plans is that the system may assist the user in the use of the available privacy plans. While it is nice to have a rich menu of tools for enhancing privacy, these tools won't have much value if the user doesn't realize when or how the tools should be used. For this reason, the system monitors the virtual world to detect situations in which a given user's privacy may be in jeopardy, and in this case the system alerts the user and suggests usage of the tools. How can a system detect when a user's privacy is in danger? Certain user-independent event patterns may come into play, for example the system might recognize that the user is within view of a crowd of avatars or bots, even if the user does not notice or cannot see these observers.

The context of user interaction is another source of information that can be used to detect when privacy is in jeopardy. For example, the system may detect that the user is about to begin a privacy-sensitive interaction with the virtual world. The user might be entering a virtual bank, for example, or may be starting to shop for sensitive items like lingerie at which point the system could recognize this pattern and suggest the usage of a privacy plan. This recommendation may be based on the location or type of interaction the user is engaged in, as well as the relative proximity of other avatars or bots. The recommendation could also take into account how the user has invoked the privacy tools in the past. For example, if the user has regularly used the Lockout Plan in the past, the system recommends Lockout when it is viable; but if such a plan is determined to be not viable due to local observers already in the space, then the system recommends using a Private Copy Plan.

Deployment

This article has focused on the logical development of privacy plans. To provide real world context, this section provides a high level view of a deployment scheme compatible with the needs of privacy plans in social (VR) metaverses. Figure 5 illustrates that playable metaverses will be served from servers over the Internet. Users (embodied as avatars in the metaverse) connect from their devices and make use of the metaverse engine and the metaverse instance (e.g., M1, M2, …) to become a part of the world. Privacy plan capabilities, such as the ability to: launch and control a plan, monitor a plan, persist a plan, and determine effectiveness of a plan, are functionally coded in a game object called Privacy Manager (PM). Such logic is loaded with the metaverse and is present in the engine on the server and on the clients. For example, if Unity3D is the basis for the metaverse, then the PM is a function that may be invoked in the scene, available through an API. In Unity each part of a metaverse is referred to as a scene, and scenes contain instances of game objects such as player and non-player avatars as well as scenery such as structures and lighting. The implementation of Privacy Plans is compatible with Unity's functional architecture.

Managing User Privacy

This article focuses on approaches and underpinnings that will help participants manage user-privacy while immersed in virtual reality worlds. A solution is needed because the social metaverse features a) large numbers of avatars, b) "open" capabilities for moving, acting, and interacting, and c) a wide and anonymous user base. These characteristics may align to enable nefarious within-metaverse tracking and surveilling. This, in turn, is likely to be annoying and even dangerous as it may compromise user privacy and personal information.

Our system provides various tools and techniques by which a VR user may preserve privacy and prevent such surveillance by others. The system may, for example:

- Provide a means for a VR user to confuse observers with noise and deceptive data.
- Allow the user to become "invisible" to other users for some period of time.
- Allow a user to inhabit a private copy of some part of the virtual world, so that the user may interact with the private copy unseen by others

These techniques will not be perfect but learning how to measure their efficacy is important. There are many other challenges for successfully deploying such tools, and our next steps are to study these in further detail. We feel that privacy mechanisms that do not preserve the continuity of the metaverse will not likely be acceptable to the users they intend to protect. Therefore, many of the challenges ahead lay at the intersection of algorithms and user experience. The battle has only just begun.

Author Information

Ben Falchuk is with Vencore Labs, NJ, U.S.A.

Shoshana Loeb is with Open Ventures L.L.C., PA, U.S.A.

Ralph Neff is with InterDigital, Inc., DE, U.S.A.

References

[1] K. Krombholz, H. Hobel, M. Huber, and E. Weippl, "Advanced social engineering attacks," *J. Information Security and Applications*, vol. 22, pp. 113-122, 2015.

[2] A. Bartoli et al., "On the ineffectiveness of today's privacy regulations for secure smart city network," *smartcitiescouncil.com*, Jun. 12, 2015; https://smartcitiescouncil.com/search-scc?keys=On+the+ineffectiveness+of+today%E2%80%99s+privacy+regulations+for+secure+smart+city+network.

[3] S. Biddle, "Sharing with "friends of friends" on Facebook exposes you to 150,000 people," *gizmodo.com*, Feb. 3, 2012: http://gizmodo.com/5882027/sharing-with-friends-of-friends-on-facebook-exposes-you-to-150000-people.

[4] J. Kopfstein, "Virtual reality allows the most detailed intimate digital surveillance yet," *The Intercept*, Dec. 23, 2016; https://theintercept.com/2016/12/23/virtual-reality-allows-the-most-detailed-intimate-digital-surveillance-yet/.

[5] E. Cline, *Ready Player One*. New York, NY: Broadway, 2011.

[6] F. Steinicke, *Being Really Virtual: Immersive Natives and the Future of Virtual Reality*. Springer, 2016.

[7] W.Y. Tang and J. Fox, "Men's harassment behavior in online video games: Personality traits and game factors," *Aggressive Behavior*, Feb. 16, 2016.

[8] C. Dewey, "Men who harass women online are quite literally losers, new study finds," *Washington Post*, Jul. 20, 2015.

[9] B.Linden, "How to deal with abuse and harassment," *Second Life Community*; https://community.secondlife.com/knowledge-base/english/how-to-deal-with-abuse-and-harassment-r610/, accessed Mar. 15, 2018.

[10] M.S. El-Seif, A. Darchen, and A. Canossa, Eds., *Game Analytics; Maximizing the Value of Player Data*. Springer, 2013.

[11] M.L. Han, J.K. Park, and H.K. Kim, "Online game bot detection in FPS game," *Aggressive Behavior*, vol. 42, no. 6, pp. 513-521, Nov, 2016.

[12] H. Kim, S. Hong, and J. Kim, "Detection of auto programs for MMORPGs," in *AI 2005: Advances in Artificial Intelligence*, S. Zhang and R. Jarvis, Eds (Lecture Notes in Computer Science). Heidelberg/Berlin: Springer, 2015.

[13] C.S. Ang, P. Zaphris, and S.Mahmood, "A model of cognitive loads in massively multiplayer online role playing games," *Interacting with Computers*, vol. 19, pp. 167-179, 2007.

[14] B. Falchuk, K.C. Lee, S. Loeb, E. Panagos, and Z. Yao, "Just-in-time reconnaissance and assistance for video game streams and players," in *Proc. IEEE Consumer Communications and Networking Conf. (CCNC'16)*, work-in-progress track, 2016.

[15] B. Falchukand and S. Loeb, "Privacy enhancements for mobile and social uses of consumer electronics," *IEEE Commun. Mag.*, vol. 48, no. 6, pp. 102-108, 2010.

[16] "Metaverse," *Wikipedia*, 2018; https://en.wikipedia.org/wiki/Metaverse; accessed Sept. 2017.

[17] M. Brian, "'Overwatch' player toxicity is delaying game updates," *engadget*, Sept. 14, 2017; https://www.engadget.com/2017/09/14/blizzard-overwatch-toxicity-developer-update/.

[18] A. Davis, "New tools to prevent harassment," *Facebook newsroom*, Dec. 19, 2017; https://newsroom.fb.com/news/2017/12/new-tools-to-prevent-harassment.

[19] E. Maiberg, "Steam is full of hate groups," *Motherboard*, Oct. 19, 2017; https://motherboard.vice.com/en_us/article/d3dzvw/steam-is-full-nazi-racist-groups.

[20] M. Moon, "Google is developing techniques to combad VR trolls," *engadget*, Aug. 10, 2016; https://www.engadget.com/2016/08/10/google-daydream-labs-vr-trolls.

[21] J. Donnelly, "The closure of OpenIV leaves Grand Theft Auto 5 machinima creators at a crossroads," *PCGamer*, Jun. 22, 1017; http://www.pcgamer.com/the-closure-of-open-iv-leaves-grand-theft-auto-5-machinima-creators-at-a-crossroads.

NIKKIJOHNSONPHOTO/ISTOCK

1932-4529/18©2018IEEE

Keeping the Lights On

A Comparison of Normal Accidents and High Reliability Organizations

Hilary Brown

Large technological systems have many modes of failure — some mundane, others exotic, some with dire consequences. Failures resulting in death and environmental degradation spring easily to mind: Chernobyl, Bhopal, Deepwater Horizon. As a large technological system, the U.S. electric power infrastructure experienced failures during major Northeast blackouts in 1965 and 2003, when large areas were left without power, and the system frequently experiences both large and small blackouts. How should we conceptualize failure in complex technological systems like electric power?

To answer this question, two frameworks have been proposed: normal accidents and high reliability organizations (HROs). The *normal accident framework* argues that accidents are endemic to complex technological systems, from analyses combining technological, social, and political concerns. Based on in-depth organizational studies, the *HRO framework* concludes that certain organizations cope with complex technological systems in ways that make failure less likely. Both frameworks emphasize complexity and technical interactions. Both examine how operators, managers, and the organizational structures interact with one another and with technological systems. Yet, they arrive at very different (though not mutually exclusive) conclusions: accidents *will* happen or accidents can be avoided.

Power systems engineers and managers strive to "keep the lights on" with technological fixes, market adjustments, and extreme work schedules to restore power quickly. However, with a few exceptions, the power systems literature has paid little attention to the frameworks of normal accidents or HROs. As a complex technological system comprising many organizations and vast amounts of equipment, the field of electric power could benefit from attention to both perspectives. First, I will examine the two frameworks individually and compare them, emphasizing any work from the literature of these fields using electric power examples or case studies. Then, I will look for evidence that either framework has been discussed from an engineering perspective in the power systems literature. Finally, I will postulate ways in which consideration of these frameworks would benefit the field of power systems.

Normal Accidents and High Reliability Organizations

First of all, what is a normal accident? It is an accident that is neither common nor expected, resulting from unforeseen interactions among system elements and proceeding in an incomprehensible way. The term came into wide use with Perrow's work, *Normal Accidents*, which classified technological systems according to two qualities: coupling and complexity [1]. A system is either tightly coupled or loosely coupled, depending on how interdependent and time-sensitive different parts of the system are with respect to one another [1]. According to Perrow's definition, the power grid is tightly coupled because electricity must be consumed and produced at the same time — it can rarely be cost-effectively stored.

Digital Object Identifier 10.1109/MTS.2018.2826059
Date of publication: 31 May 2018

The coupling of the power grid is loosened by the substitutability of generation: during times of low demand, the electricity generated from hydropower or from coal is equivalent. Interactions in a technological system may be either linear and therefore more easily comprehensible and predictable, or nonlinear and complex, in which seemingly disconnected parts of the system affect one another [1]. Abrupt changes in generation and load in one part of the power system affect the frequency of the whole system — a disturbance in Atlanta, GA, could appear in frequency measurements in Albany, NY. Normal accidents occur in tightly-coupled systems with complex interactions.

Perrow studied nuclear power plants, chemical processing facilities, and air traffic control to characterize normal accidents [1]. After examining accident reports blaming "operator error," Perrow cautions that "human error" is a convenient catch-all for inexplicable accidents and using such phrases may indicate a normal accident has occurred [1]. After failures, it is natural to fall back on the common engineering technique of adding redundancy to ensure reliability. However, Perrow argues that this adds to interactive complexity and actually exacerbates the potential for normal accidents [1]. He also notes that despite a veneer of safety-consciousness by companies, production pressures usually supersede safety considerations and operators are rarely able to protest [1]. Normal accidents are less likely in systems with strong employee representation, according to Perrow [1].

In a later article, Perrow identifies characteristics of "error-avoiding" systems. First, error-avoiding systems are experienced with the scale of operation and activity during the critical phase [2]. Error-avoiding systems also collect information on errors and interact with elites; for example, CEOs and Congresspeople make extensive use of air travel [2]. On an organizational level, error-avoiding systems exert control over their members and the system environment is dense, with many different firms, regulators, and interest groups [2]. Perrow argues that the system becomes "error-neutral" or "error-inducing" if any of these elements are absent [2].

The postscript to the 1999 edition of *Normal Accidents* discusses the electric power system and the looming Y2K crisis. Perrow explains that "Y2K could be the quintessential Normal Accident… small failures to read the correct date cannot be fully anticipated, and their interactions cannot be imagined; the tight coupling of our highly interdependent systems can bring about a cascade of failures" [1]. Perrow criticizes the optimism of a North American Electric Reliability Corporation (NERC) report asserting the industry was "ready" to deal with Y2K, in the sense of "coping with" rather than avoiding problems [1]. Perrow further disagreed with the report's emphasis on continuing deregulation activities instead of addressing Y2K problems [1]. In hindsight, the Y2K problem was successfully remediated, despite predictions to the contrary.

Ultimately, Perrow uses the framework of normal accidents to argue that complex technological systems can never be completely safe. He argues that hazardous technological systems with unborn victims, like nuclear power, should not be used because the inevitable failure far outweighs any economic gains.

Proponents of the HRO framework, on the other hand, believe certain organizations experience failures less often than expected and that technological systems can be managed more-or-less safely. After a major accident occurs, Roberts notes that solutions may include technological fixes or bans, but typically less attention is devoted to improving the system management [3]. HRO researchers explore why certain organizations have higher reliability than others; they focus on organizations that have already achieved high levels of reliability and work backwards to determine how. Early HRO case studies focused on naval aircraft carriers, air traffic control, grid operations, and a nuclear power plant [3]–[7].

HRO researchers argue that the framework addresses a gap in organizational theories based on dichotomies: planning versus trial-and-error, certainty versus uncertainty, and hierarchy versus decentralization [4]. LaPorte and Consolini believe that HROs, although hierarchical and planning-oriented, bridge these categories by allowing flexibility and decentralized decision making in certain situations [4]. HROs use technologies that are tightly coupled to the organization (technological failure threatens organizational failure), have a strong external preference for failure-free operations, and invest heavily in reliability improvements [4]. For all HROs, the cost of failure is much higher than the value of the lessons learned [4].

HROs develop reliability through redundancy, frequent training, emphasizing responsibility, and distributing decision-making throughout the group hierarchy, all of which reduce the impacts of complexity and tight coupling, as defined by Perrow [3]–[6], [8]. For example, continuous training allows operators to gain experience with novel system behaviors and proficiency with complex technology [5]. Redundancy reduces the impact of tight coupling by responding to time-dependent processes and creating multiple paths for success [5]. HROs try to replace indirect information sources (complexity) with many direct information sources [5]. Through all these actions, HROs develop a "culture of reliability," described by Roberts as including interpersonal responsibility, person-centeredness, strong feelings of credibility, an emphasis on creativity, and being helpful and supportive [3], [6]. Because using total failure to learn is not possible, HROs complete immediate investigations of small incidents and quickly discuss lessons learned [4].

Through their interactions, the people in an HRO create a "collective mind," a concept to describe the ways in which the conscious attention of many individuals is linked together into a structure of social cognition [7]. This increases reliability in three ways [7]: 1) connecting across time by bringing knowledge forward from previous parts of a process, 2) incorporating more and more tasks into the framework of cognition, and 3) connecting new and old employees, allowing bi-directional learning, as experienced employees interact with new employees who see the system with a fresh perspective. Using "collective mind," Weick and Roberts argue that a complex cognitive structure is the only way to comprehend the complexity of a large technological system, since no individual can grasp it [7].

Some of the first studies to develop the concept of HROs examined power systems, specifically that of Pacific Gas and Electric. In [5], Roberts examines how training helps operators cope with varied system conditions. Grid operators at Pacific Gas and Electric train using recent issues from actual operations, while the nuclear plant operators spend one week per month training, allowing them to stay up-to-date with unique and/or dangerous conditions [5]. At Pacific Gas and Electric, Roberts noticed frequent training, redundancy, and distributed decision-making, which were then used to generally describe HROs.

About ten years later, during the electricity market restructuring in California, HRO researchers revisited Pacific Gas and Electric. They witnessed what is now commonly called the California Electricity (or Energy) Crisis. Schulman and his co-authors pose the following question in their studies: how do technical systems with many players maintain reliability [9]? They argue that electricity restructuring is a good test case for two reasons [9]: 1) reliability should be undermined based on early research, and 2) it challenges whether complexity and tight coupling cause failure. The authors argue that, despite rolling blackouts and the declaration of bankruptcy by Pacific Gas and Electric, "the lights by and large actually stayed on — and *reliably* stayed on" [9].

I disagree with the authors' assessment of this point. The utilities and operators struggled valiantly during the crisis and achieved impressive results. However, Schulman *et al.* understate the impacts. Weare explains that:

> ■ [t]he lights flickered throughout the crisis… In 2000, electricity was turned off to customers with special interruptible contracts on 13 other days. During 2001, "load shedding" occurred on 31 days. On nine of these days customers experienced involuntary rolling blackouts for a total of 42 hours of outages. During these nine outages, California experienced an average shortfall of… enough energy to power over 450,000 households. On the worst day, January 18, the equivalent of almost one million households lost electricity [10].

High Reliability Organizations can be designed to compensate for human fallibility and technological failures.

This description does not meet the basic expectations of "high reliability."

Schulman and his co-authors coin the phrase "high reliability network" to describe the constellation of utilities, generation owners, and the independent system operator [9]. Unlike HROs, the authors show that high reliability networks engage in trial-and-error, based on reliability standards and actual operating conditions, and follow rules of thumb to ensure that the system's performance remains within the specified requirements [9], [11]. Roe writes that "[r]eliability here lies in *resilience* — the ability of managers to respond in ways that buffer or accept variance in inputs and then act to counter the variance in order to produce output fluctuations at manageable levels" (emphasis added) [11]. Unfortunately, reliability and resilience, though related, are not the same. (Reference [12] has an extended discussion on different definitions of resilience.)

In summary, the HRO framework claims that organizations can be designed to compensate for human fallibility and technological failures. The normal accident framework claims that reliability cannot be guaranteed due to fundamental system characteristics. These frameworks are frequently compared to one another, since both examine failures in complex systems. Rosa summarizes three major contradictions between the HRO framework and the normal accident framework. The latter expects infrequent, but "normal," accidents while the former expects virtually accident-free operation [13]. Both frameworks predict some accidents; the question is how to define "infrequent," which neither framework specifies [13]. The normal accident framework posits that redundancy is a cause of failure, while the HRO framework argues that redundancy decreases accidents [13]. The normal accident framework rests on asymmetrical social and political power structures and the HRO framework focuses on a culture of reliability [13]. To Rosa, the two frameworks are "blindfolded observers feeling different parts of an elephant" [13], where the elephant represents complete understanding of failures in complex technological systems.

In his 1993 book, Sagan uses both frameworks to examine nuclear missile defense. He argues that, superficially, the nuclear missile defense system seems to support the HRO framework since accidental nuclear war has not yet occurred [14]. Upon detailed inspection, Sagan determined that several incidents seemed more consistent with normal accidents. Instead of learning from mistakes, as predicted by the HRO framework, the people involved in these incidents attempted to simultaneously cover them up and spin the facts to support the continued development of weapons systems [14]. Furthermore, he identifies challenges to organizational learning cited by the HRO framework: ambiguous feedback, political considerations, accurate reporting, and secrecy. These "other" concerns dominated safety considerations after incidents, according to Sagan, and prevented military organizations from learning from failures of the nuclear missile system.

Sagan's characterization of the HRO framework inspired a debate in the *Journal of Contingencies and Crisis Management*, including articles from LaPorte, Perrow, and Sagan [2], [8], [15], [16]. LaPorte disagreed with Sagan's assessment of HROs as "optimistic" and agreed with Perrow that avoiding failure cannot be guaranteed. Perrow's response accused the HRO researchers of failing to critically engage with the organizations that they observe and wrote that "no one can be against clear safety goals, learning, experience, and so on" [2]. Perrow argued that Sagan's greatest contribution was to emphasize the role played by group interests in accidents. In response, LaPorte and Rochlin asserted that the HRO framework is a study of organizations under trying conditions — not a theory of accidents — and to directly compare it to the normal accident framework is fruitless [15]. Finally, Sagan closed the debate by calling for focus on the political aspects of accidents and organizations [16]. He specifically argued for more research on redundancy [16]: when it creates common-mode failures, when it decreases component reliability, and when organizational redundancy is equivalent to engineering redundancy (argued by Roberts in [6]).

To summarize, HRO researchers see a complementary role for each framework, while normal accident researchers tend to see little of interest in the in-depth organizational studies supporting HRO research. We have seen from this discussion that HRO researchers focus on the organizational structure of the organization, while normal accident researchers focus on the underlying characteristics of the technological system. We cannot ignore the underlying technology, but we also need to know the role that organizations play in enabling or avoiding failures. Normal accident researchers emphasize political and social power relationships, while HRO researchers emphasize cooperation. Both frameworks are useful, because they explore different, yet important, elements of human behavior and complex technological systems.

Electric Power Systems as Complex Technological Systems

The power grid has always been a large aggregate of equipment and organizations, as described by Hughes [17], [18]. In the U.S., there are more than 19 000 individual generators rated larger than 1 MW in more than 7000 power plant facilities [19]. The high voltage transmission system comprises more than 640 000 miles of lines, while the distribution system has more than 6.3 million miles of lines [19]. On the organizational side, deregulation created a plethora of players by dismantling many vertically-integrated monopolies and opening opportunities for competition. As a complex system regulated for reliability, electric power systems offer an interesting case study for both the normal accident framework and the HRO framework. As discussed in the previous section, both Perrow and HRO researchers (Roberts and Schulman) have devoted attention to electric power systems. Now, the ways both frameworks have been taken up in electric power systems literature will be discussed.

In 2001, a workshop on critical infrastructure and interdisciplinary research convened in Washington, DC. In the first session, Perrow admitted that his fears about Y2K did not come to pass, explaining that the world was "less interactively complex and tightly coupled than some of us... thought it would be" [20]. Perrow is concerned that increased centralization through mergers and/or market control is being used to deal with complex interdependencies [20], though, in previous work, he argued that only organizations with rigid hierarchy and strong discipline could deal with complex and tightly coupled systems [1]. In his presentation, Perrow focused on the need for collaboration because limited communication and increased centralization will lead to more failures as the world becomes more complex and tightly coupled [20].

Peerenboom described different types of interdependency and failure, giving more explanatory power to interdependency (coupling), something he previously described in [21]. Interdependency can be physical, where the material output of one infrastructure system is used by another [22], such as using electricity to extract coal to generate electricity. Interdependencies can also be "cyber" (electronic information and control systems) or geographic (infrastructure is co-located without connections) [22]. Finally, Peerenboom defines logical interdependency to capture other coupling, e.g. financial markets [22].

Peerenboom distinguishes between cascading, escalating, and common cause failures between infrastructures. In his representation, cascading failures are those in which a disruption in one infrastructure causes failures in another, while escalating failures are those for which a disruption in one infrastructure exacerbates an independent disruption in another [22]. This differs from the more narrow definition of a cascading failure within the electric power system research community: "a sequence of dependent failures of individual components that successively weakens the power system" [23]. Common cause failures occur when two or more infrastructures are disrupted simultaneously, by a severe storm for example [22]. Peerenboom, like Perrow, called for cross-disciplinary collaboration, a call continued in [24], which apparently includes those who study normal accidents or HROs through the generic qualifier of "social scientists."

A 2004 opinion piece argued that the electric power system has a lot to learn from air traffic control [25]; one of the authors participated in the aforementioned critical infrastructure workshop [26]. Air traffic control is a darling of both frameworks. From the normal accident framework, the authors recognize that operations and investigations into failure must be located in separate agencies [25], presumably to reduce political and social pressures. They also propose that national coordination is needed, although local and regional actions dominate [25]. Consistent with the HRO framework, the authors suggest transforming panicked responses when things go wrong into incident investigations and research and development for new tools [25]. The authors also note the need for comprehensive data monitoring and real-time interpretation [25], helping to create a big-picture view of the system. Although the authors do not mention either framework, it is clear that their suggestions for changes in the electric power system were inspired by case studies and findings from both frameworks.

In a 2006 conference paper, Hines *et al.* examined large blackouts in the United States to test the assumption that blackout frequency (adjusted for demand growth) should decrease due to engineering and policy changes made after each large blackout since the late 1960s [27]. The authors argue that "[t]he U.S. air traffic control system provides precedent for a large, complex system undergoing a significant decrease in risk following appropriate engineering and policy actions" [27]. Using data from 1984 to 2000, the authors note that "human error" accounts for 11% of all reported disturbances [27]. (Recall Perrow's warning about this label). Excluding outages caused by weather, the authors found that the frequency of large blackouts is not decreasing, speculating that the possible increase is caused by under-investment in transmission and a lack of mandatory, enforceable rules for reliability [27]. They also note that the protection system, which isolates stressed equipment to avoid damage to the individual component, tends to cause cascading failures at the system level rather than control them [27].

When systems have self-organizing criticality, actions designed to mitigate problems may actually increase the likelihood of large disruptions.

In a 2009 journal article, the same group of authors found that blackout frequency: 1) has not decreased with time, 2) changes seasonally, and 3) increases during times of peak use [28]. Furthermore, the size of the blackout follows a power law probability distribution (as shown by other researchers) and is not correlated with restoration time [28]. They recommend doubling the operators during peak use times and focusing mitigation on both large and small outages [28]. Although Hines *et al.* do not acknowledge either normal accidents or HROs, their findings and policy recommendations could have benefited from such attention. Whether organizational redundancy (adding more operators during peak use times) will actually help is an open question. HRO researchers believe that such redundancy is important, while normal accident researchers are unconvinced. A footnote seems to acknowledge this, noting that "most electric utilities, and other system operators, increase operations staff during peak periods and daytime hours" [28]. Why has the blackout risk during peak hours persisted even though utilities use increased organizational redundancy? The normal accident framework would suggest that it is because the electric grid is more tightly coupled as load increases. The HRO framework might suggest ways in which organizational redundancy is not actually fulfilling its desired function through "mindlessness" or limited cooperation [29].

In 2011, a group of power system researchers investigated the "complex systems aspects of blackout risk and mitigation" [30]. By focusing on the frequency distribution of different sized blackouts, the authors argue that components cannot be assumed to be independent [30]. If independence was a valid assumption, the probability of large blackouts would show an exponential decay; instead, it follows a power law distribution [30], [31]. The authors claim that this property of the blackout frequency distribution supports Perrow's description of interactive complexity and that large disruptions are seemingly intrinsic to large infrastructure systems [30]. They then describe self-organizing criticality as a property of complex systems, where the nonlinear system dynamics in the presence of perturbations actually make the average system state more susceptible to large disturbances [30]. When systems have self-organizing criticality, actions designed to mitigate problems may actually increase the likelihood of large disruptions [30]. This paper reinforced ideas central to the normal accident framework with engineering theory, such as independence and self-organizing criticality.

The same authors (Dobson, Carreras, and Newman) and their collaborators (Lynch and Ren) continued exploring the probability distribution of blackouts in [32]–[34], focusing on the influence of different network assumptions. These papers used engineering analysis to answer questions relevant to both frameworks: how do redundancy and scale change the likelihood of blackouts? The researchers model the evolution of the power system considering policy (system upgrades) and societal changes (load growth and the corresponding power supply growth) [32]–[34]. This model demonstrates how the power system self-organizes to its critical point — serving increased load stresses the system and system upgrades responding to that stress, in turn, allow for more load to be served, keeping the network near its critical point [32], [33].

The researchers examined three different approaches to line upgrades to evaluate their impact on blackout probability: 1) upgrade as lines approach their loading limits [32], 2) upgrade lines involved in a cascading outage after the outage [32], [33], and 3) upgrade when lines violate the N-1 criterion [33]. The power grid is designed to operate so it satisfies the "N-1" criterion, meaning that the grid should remain fully operational when any one major piece of equipment suffers an outage. To satisfy this criterion, lines cannot be loaded to capacity; it provides a pseudo-redundancy, with spare capacity shared over multiple lines. Comparing the first two approaches, no difference in the likelihood of large blackouts was found [32]. The second approach showed greater grid utilization, while the third reduced the number of small outages [33]. None of the approaches yielded a lower probability of large blackouts than the others [32], [33]. Furthermore, the researchers found that increasing the lines' reliability (mathematically equivalent to decreasing the margin), actually increased the probability of larger blackouts [32]. Defining outage risk using probability and cost, the risk of large blackouts was shown to increase with grid size [34]. This paper is noteworthy because it tested the normal accident framework against the engineering heuristic that larger grids reduce outages, concluding that bigger is better only until a threshold at which the risk of smaller blackouts is balanced by that of larger blackouts [34].

In their 2012 article, Mazur and Metcalfe analyzed three different power grids in the U.S. to determine whether grid size determines reliability [35]. Proponents of the normal accident framework would expect benefits from increased integration to be outweighed by the increased risk of catastrophic failure. The authors ultimately found no relationship between grid size and reliability, finding that the normal accident framework does not accurately predict performance of the electric power system [35]. However, their study only examined data from 2007 to 2010, excluding the 2003 Blackout. Normal accidents happen "infrequently," with relevant time intervals typically measured in decades, so studies using only three years of reliability data cannot make claims either for or against the framework. Based on 22 years of historical outage data, Dobson, Carreras, and Newman characterized the probability of blackouts for the Eastern and Western interconnections [36]. The study results used historic data to confirm the power law distribution mentioned previously [30], supporting the normal accident framework.

Despite the clear applicability of both the normal accident framework and the HRO framework, relatively few groups of power systems engineers have adopted ideas from either. As described, most work uses either framework to analyze blackouts, specifically large or cascading blackouts. I will now postulate on some specific areas where both frameworks could be applied and discuss how the results from or questions inspired by either framework would be useful. I will focus on cascading blackouts, digital relays for protection and control, and cybersecurity.

In the post-event analysis of the U.S. 2003 Northeast Blackout, investigators traced paths of failure that were unclear to operators during the event, who instead saw baffling interactions and an incomprehensible evolution of problems — hallmarks of normal accidents. The final report on the blackout cited "inadequate system understanding … situational awareness … [and] diagnostic support" as causes of the cascading blackout [37]. It also said that "[m]any of the institutional problems arise not because NERC is an inadequate or ineffective organization, but rather because it has no structural independence from the industry it represents and has no authority to develop strong reliability standards and to enforce compliance" [37]. This recalls Perrow's argument that the regulator must be independent of the industry it regulates to reduce normal accidents. Engineers can examine the role that redundancy played: did it contribute to the blackout or did it help keep the grid energized? The HRO framework could help companies learn about management strategies — which companies responded well to the unfolding events and which exacerbated the problem? The report cited above focuses on violations of voluntary reliability standards rather than delving into the company culture(s) responsible for those violations [37].

Digital relays for protection and control offer another fascinating comparison of normal accidents and HROs. In substation design, most protection and control systems are fully redundant. Yet, relay misoperations persist and are most commonly caused by 1) incorrect settings, logic, or design error, 2) relay failure or malfunction, and 3) communication failure [38]. The dominant relaying philosophy is to isolate stressed equipment and protect it from damage. The authors in [27] believe relays should be set considering the cost of an outage to the *system*, rather than only the cost of the protected *component*. Considering the system implications of isolating equipment may allow transmission lines to be overloaded, but prevent a cascading outage. Are group interests at play in protecting equipment at the expense of the system?

The authors in [39] argue that the most troublesome relay misoperations are "hidden failures" and believe that changing relays from an "OR" selection to a voting system would help reduce misoperations [39]. However, with misoperations spread between design, equipment, and communication, it is fair to say that the fundamental reasons for persistent misoperations have not yet been identified. Is the problem in the protection philosophy, as argued in [39]? Or is it with the management and engineering design of the system? The normal accident framework could help identify whether tradeoffs between redundancy and complexity are causing problems in this application. The HRO framework could help to identify management issues and offer advice on how to change.

In the arena of cybersecurity, the number of vulnerabilities is countless [40], partially due to complexity and tight coupling. Even so, many vulnerabilities are known [41]–[43] and cyberattacks are already common in power systems [44]. How much effort should be taken to protect the power system from these? Some researchers promote "resiliency" as the best way to deal with cybersecurity issues. Resiliency is nothing but the ability to bounce back from failure(s) quickly. Can the HRO framework teach us something about resiliency and how organizational design can promote it? Normal accident researchers may suggest how to manage the political and social issues surrounding cybersecurity violations.

Making Progress in Understanding Failure

Overall, the power systems literature has relatively few mentions of insights to be gained from either the normal accident or HRO frameworks. This absence indicates an area where progress could be made in understanding failures. Studies of cascading blackouts illustrate ways in which both frameworks would aid analysis. Similar studies of other power systems topics, like digital relays and cybersecurity, could also benefit from attention to both frameworks. From the normal accident framework, engineers can determine the point where benefits from redundancy are offset by increased complexity and closely examine hierarchical power structures and market pressures to determine whether any of these elements are undermining their quest for reliability. From the HRO framework, engineers can acknowledge that sometimes the fix is not a fancy new technology, but rather changing organizational culture. None of this will be easy, but if we want to have trustworthy complex technological systems, then we must know why they fail and how (or whether) those failures can be avoided.

If we want trustworthy complex technological systems, then we must know why they fail and how (or whether) those failures can be avoided.

Acknowledgment

This research was supported by the U.S. Department of Energy Lawrence Berkeley National Laboratory, through the project "Algorithmic/Computational Research within Power System Vulnerability," under a prime contract No. DE-AC02- 05CH11231.

Author Information

Hilary Brown was with the University of Wisconsin-Madison, where she received her Ph.D. degree in electrical engineering, with a doctoral minor from the Holtz Center for Science and Technology Studies. She now works in transmission planning in Minnesota. Email: brown .hilary@ieee.org.

References

[1] C. Perrow, *Normal Accidents: Living with High-Risk Technologies*. Princeton, NJ: Princeton Univ. Press, 1999.

[2] C. Perrow, "The limits of safety: The enhancement of a theory of accidents," *J. Contingencies and Crisis Management*, vol. 2, pp. 212-220, Dec. 1994.

[3] K. H. Roberts, "Cultural characteristics of reliability enhancing organizations," *J. Managerial Issues*, vol. 5, pp. 165-181, 1993.

[4] T.R. LaPorte and P.M. Consolini, "Working in practice but not in theory: Theoretical challenges of high-reliability organizations," *J. Public Administration Research and Theory*, vol. 1, pp. 19-48, 1991.

[5] K. H. Roberts, "Managing high reliability organizations," *California Management Rev.* vol. 33, pp. 101-113, Sum. 1990.

[6] K. H. Roberts, "Some characteristics of one type of high reliability organization," *Organization Science*, vol. 1, pp. 160-167, 1990.

[7] K. E. Weick and K. H. Roberts, "Collective mind in organizations: Heedful interrelating on flight decks," *Administrative Science Quart.*, vol. 38, pp. 357-381, Sept. 1993.

[8] T. R. LaPorte, "A strawman speaks up: Comment on *The Limits of Safety*," *J. Contingencies and Crisis Management*, vol. 2, Dec. 1994.

[9] P. Schulman, E. Roe, M. van Eeten, and M. de Bruijne, "High reliability and the management of critical infrastructures," *J. Contingencies and Crisis Management*, vol. 12, pp. 14-28, 2004.

[10] C. Weare, "The California electricity crisis: Causes and policy options," Public Policy Institute of California, San Francisco, CA, 2003. [Online]. Available: http://www.ppic.org/content/pubs/report/R_103CWR.pdf. [Accessed: May 14, 2015].

[11] E. Roe, P. Schulman, M. van Eeten, and M. de Bruijne, "High-reliability bandwidth management in large technical systems: Findings and implications of two case studies," *J. Public Administration Research and Theory*, vol. 15, pp. 263-280, 2005.

[12] S. Tas, "A comprehensive method for assessing the resilience of power networks in the face of an intelligent adversary," Ph.D. dissertation, Dept. of Ind. Eng., Univ. of Wisconsin-Madison, Madison, WI, U.S.A., 2012.

[13] S.D. Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton Univ. Press, 1993.

[14] E.A. Rosa, "Celebrating a citation classic – and more: Symposium on Charles Perrow's *Normal Accidents*," *Organization and Environment*, vol. 18, pp. 229-234, Jun. 2005.

[15] T.R. LaPorte and G. Rochlin, "A rejoinder to Perrow," *J. Contingencies and Crisis Management*, vol. 2, pp. 221-227, Dec. 1994.

[16] S.D. Sagan, "Toward a *political* theory of organizational reliability," *J. Contingencies and Crisis Management*, vol. 2, pp. 228-240, Dec. 1994.

[17] T.P. Hughes, *Networks of Power: Electrification in Western Society, 1880-1930*. Baltimore, MD: Johns Hopkins Univ. Press, 1983.

[18] T.P. Hughes, *American Genesis: A Century of Invention and Technological Enthusiasm, 1870-1970*. Chicago, IL: Univ. of Chicago Press, 2004.

[19] "Quadrennial Energy Review: Energy Transmission, Storage, and Distribution Infrastructure," Dept. of Energy, Washington, DC, Apr. 2015. [Online]. Available: http://energy.gov/epsa/quadrennial-energy-review-qer. [Accessed: Apr. 30, 2015].

[20] C. Perrow, "A university perspective on critical infrastructures," presented at *Workshop on Critical Infrastructure: Needs in Interdisciplinary Research and Graduate Training*, C.L. DeMarco, Ed. Washington, DC, 2001.

[21] S.M. Rinaldi, J.P. Peerenboom, and T.K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Systems Mag.*, pp. 11-25, vol. 21, no. 6, Dec. 2001.

[22] J. Peerenboom, "Infrastructure interdependencies: Overview of concepts and terminology," presented at *Workshop on Critical Infrastructure: Needs in Interdisciplinary Research and Graduate Training*, C.L. DeMarco, Ed. Washington, DC, 2001.

[23] IEEE PES CAMS Task Force on Understanding, Mitigation and Restoration of Cascading Failures, "Initial review of methods for cascading failure analysis in electric power transmission systems," presented at the IEEE Power and Energy Soc. General Meeting, Pittsburgh, PA, U.S.A., Jul. 20-24, 2008.

[24] J.P. Peerenboom and R.E. Fisher, "Analyzing cross-sector interdependencies," in *Proc. 40th Hawaii Int. Conf. System Sciences*, Jan. 3-6, 2007 (Waikoloa, HI). [Online]. Available: IEEE Xplore, http://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: May 11, 2015].

[25] J. Apt, L.B. Lave, M.G. Morgan, M. Ilic, and S. Talukdar, "Electrical blackouts: A systemic problem," *Issues in Science and Technology*, Sum. 2004. [Online]. Available: http://issues.org/20-4/apt/. [Accessed: May 6, 2015]

[26] M. Ilic, "Change of paradigms in complexity and interdependencies of infrastructures: The case for flexible new protocols," presented at *Workshop on Critical Infrastructure: Needs in interdisciplinary research and graduate training*, C.L. DeMarco, Ed. Washington, DC, 2001.

[27] P. Hines, J. Apt, H. Liao, and S. Talukdar, "The frequency of large blackouts in the United States electrical transmission system: An empirical study," presented at the *Carnegie Mellon Conference in Electric Power Systems: Monitoring, Sensing, Software, and Its Valuation for the Changing Electric Power Industry, Jan. 11-12, 2006, Pittsburgh, PA.* [Online]. Available: https://www.ece.cmu.edu/~electricityconference/2006/hines_blackout_frequencies_final.pdf. [Accessed: May 10, 2015].

[28] P. Hines, J. Apt, and S. Talukdar, "Large blackouts in North America: Historical trends and policy implications," *Energy Policy*, vol. 37, pp. 5249-5259, 2009.

[29] K.E. Weick and K.M. Sutcliffe, *Managing the Unexpected: Assuring high performance in an age of complexity*. San Francisco, CA: Jossey-Bass, 2001.

[30] D.E. Newman, B.A. Carreras, V.E. Lynch, and I. Dobson, "Exploring complex systems aspects of blackout risk and mitigation," *IEEE Trans. on Reliability*, vol. 60, pp. 134-143, Mar. 2011.

[31] I. Dobson, "Where is the edge for cascading failure?: Challenges and opportunities for quantifying blackout risk," presented at the *IEEE Power Engineering Society General Meeting, Jun. 24-28, 2007 Tampa, FL*. [Online]. Available: IEEE Xplore, http://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: Sept. 7, 2015].

[32] D.E. Newman, B.A. Carreras, V.E. Lynch, and I. Dobson, "Evaluating the effect of upgrade, control and development strategies on robustness and failure risk of the power transmission grid," in *Proc. of the 41st Annu. Hawaii Int. Conf. on Syst. Sci.*, Waikoloa, HI, U.S.A., Jan. 7-10, 2008.

[33] H. Ren, I. Dobson, and B.A. Carreras, "Long-term effect of the N-1 criterion on cascading line outages in an evolving power transmission grid," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1217-1225, Aug. 2008.

[34] B.A. Carreras, D.E. Newman, and I. Dobson, "Does size matter?," *Chaos*, vol. 24, no. 2, 2014.

[35] I. Dobson, "Where is the edge for cascading failure?: Challenges and opportunities for quantifying blackout risk," presented at the *IEEE Power Engineering Society General Meeting, Jun. 24-28, 2007, Tampa, FL*. [Online]. Available: IEEE Xplore, http://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: May 11, 2015].

[36] A. Mazur and T. Metcalfe, "America's three electric grids: Are efficiency and reliability function of grid size?," *Electric Power Systems Res.*, vol. 89, pp. 191-195, 2012.

[37] "Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations," U.S.-Canada Power System Outage Task Force, Apr. 2004. [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf. [Accessed: May. 14, 2015].

[38] "Misoperations report," Protection System Misoperations Task Force, North American Electric Reliability Corporation, Atlanta, GA, Apr. 1 2013. [Online]. Available: http://www.nerc.com/docs/pc/psmtf/PSMTF_Report.pdf. [Accessed: May. 14, 2015].

[39] J. De La Ree, Y. Liu, L. Mili, A. Phadke, and L. Da Silva, "Catastrophic failures in power systems: Causes, analyses, and countermeasures," *Proc. IEEE*, vol. 93, pp. 956-964, May 2005.

[40] B. Schneier, *Secrets and Lies: Digital Security in a Networked World*. New York, NY: Wiley, 2000.

[41] P.K. Kerr, J. Rollins, and C.A. Theohary, "The Stuxnet computer worm: Harbinger of an emerging warfare capability," Congressional Research Services, Washington, DC, 2010. [Online]. Available: https://www.fas.org/sgp/crs/natsec/R41524.pdf. [Accessed: May 1, 2015].

[42] "Alert ICS-ALERT-14-281-01B: Ongoing sophisticated malware campaign compromising ICS (update B)," U. S. Dept. of Homeland Security, Washington, DC, Dec. 10, 2014. [Online]. Available: https://ics-cert.us-cert.gov/alerts/ICS-ALERT-14-281-01B. [Accessed: May 2, 2015].

[43] P.F. Roberts, "If cyberwar erupts, America's electric grid is a prime target," *The Christian Science Monitor*, Dec. 23, 2014. [Online]. Available: http://www.csmonitor.com/World/Passcode/2014/1223/If-cyberwar-erupts-America-s-electric-grid-is-a-prime-target. [Accessed: May 2, 2015].

[44] S. Baker, N. Filipiak, and K. Timlin, "In the dark: Crucial industries confront cyberattacks," Center for Strategic and International Studies, Washington, DC, 2011. [Online]. Available: http://www.mcafee.com/us/resources/reports/rp-critical-infrastructure-protection.pdf. [Accessed: Apr. 30, 2015].

Smart IoT Devices in the Home

Security and Privacy Implications

Vijay Sivaraman, Hassan Habibi Gharakheili, Clinton Fernandes, Narelle Clark, and Tanya Karliychuk

BEEBRIGHT/ISTOCK

Internet of Things (IoT) devices possess network capabilities and contain at least a part of the application logic, i.e., they have the ability to perform Transmission Control Protocol/Internet Protocol (TCP/IP) communications on their own, and can process some of the sensor data. The IoT thus refers to the network of physical objects embedded with electronics, software, sensors and connectivity to enable objects to exchange data with the manufacturer, operator, and/or other connected devices. At the start of this decade, there were an estimated 12.5 billion IoT devices, almost twice as much as the world's population of 6.8 billion people [1]. The number of IoT devices is expected to grow rapidly in coming years.

These technological changes have tremendous implications for decentralized production control in manufacturing, and are expected to trigger a fourth industrial revolution, following the steam engine, the conveyor belt, and the computer revolution. IoT devices will have a transformational effect on the lives of everyday consumers, too. Australia's largest telecommunications company, Telstra, says the average Australian household in 2017 had 13 Internet connected devices and that by 2021 a typical home will have over 30. It's predicted that the collective value of the smart home market in Australia will be greater than AU$1billion annually by 2021 [2]. As the IoT technology becomes embedded in televisions, webcams, smoke alarms, fitness trackers, climate-control systems, lightbulbs and more, it has the potential to save money and time, help people stay fit, healthy, and safe, and enable effortless communication with friends and family. There are important security

Digital Object Identifier 10.1109/MTS.2018.2826079
Date of publication: 31 May 2018

and privacy implications for consumers [3], however; many Internet-connected devices have poor in-built security measures [4] and can reveal private data and information that may harm or embarrass consumers [5]. A 2015 inquiry into data retention by the Parliamentary Joint Committee on Intelligence and Security (PJCIS) [6] mentioned "privacy" nearly 400 times. It said that privacy and security concerns "are closely related, as the potential for security breaches has significant ramifications for the proportionality and privacy risks associated with the proposed scheme."

IoT Consumer Research: Scenario, Test, Evaluate, Propose

In this article, we examine the security and privacy implications of selected IoT devices, building on previous work [7] in this area. Our specific contributions are as follows: First, we developed hypothetical scenarios of household IoT usage. We then tested the security and privacy vulnerabilities of several of these devices, subjecting them to hostile targeting under laboratory conditions. Next, we invited IoT suppliers, consumers, insurers, and regulators to evaluate our results at a workshop. Finally, after examining their reactions and discussing their expectations, we proposed possible approaches to help mitigate the identified risks. We also identified a research trajectory that would begin a new four-step cycle of Scenario-Test-Evaluate-Propose. We wish to emphasize that the workshop phase of our research cycle is as critical as the other phases, and not merely an afterthought. It is this phase that enables us to engage with consumers and understand the contexts in which they use their devices. In doing so, we are in a better position to construct realistic scenarios to guide our laboratory testing.

Scenarios

We created four scenarios in which people are likely to use IoT devices. Our aim was to identify products they would purchase so that we could evaluate their vulnerability under laboratory conditions. All the characters and locations are fictitious, but the scenarios are extremely realistic, and constructed on the basis of direct engagement with consumer advocates.

In the first scenario, the consumer is Tuan, a mid-career private investigator who lives by herself in a regional town in Australia, regularly drives to Melbourne and flies to Sydney to meet with clients. Most of her work involves insurance fraud although she is often asked to track cheating spouses. Because she travels quite a bit, and meets a lot of unusual people in her line of work, Tuan is worried about leaving her home unattended. Knowing the benefits of surveillance tools, she believes that installing IoT devices would offer some peace of mind. As a sole occupier who desires home security, Tuan buys three IoT devices:

1) a Belkin motion sensor to detect movements inside her house;
2) TP-Link indoor and outdoor motion sensor cameras; and
3) A Nest smoke alarm to send alerts to her smartphone in case of fire.

In the second scenario, the IoT device users are Joe and Lorna Jones, an elderly couple who live in the inner city. Lorna is a bit hard of hearing, wears a pacemaker, and has respiratory difficulties. She is not a regular user of the Internet. Joe has some mobility problems and relies on his medical-alert device when he's away from home. Lorna was playing bowls (lawn bowling) the last time he had a fall, and it took hours before he could get help. Their son, Geoffrey, who lives with his family on the Gold Coast 100-km away, wants a way to monitor his parents' welfare more thoroughly than checking in on Skype every couple of days. He has installed a number of IoT devices in their home to allow him to keep a virtual eye on Joe and Lorna's health and wellbeing. These devices are:

1) Blipcare blood pressure monitor, which sends readings to the web for Geoffrey to check;
2) Withings weighing scale;
3) Withings sleep monitor;
4) Awair air quality monitor; and
5) Netatmo weather station.

In the third scenario, Suresh and Veda Singh live in Sydney's suburbs. They know they have to cool their west-facing house in summer. Although they've trained their three growing children to moderate their electricity usage, it still feels like they're in a losing battle against the large electricity bill that arrives every quarter. While shopping for smart devices intended for use around the home, they also bought an interactive doll for their youngest child. The cute doll has a microphone that "listens" to the child, and replies in a manner similar to Apple's Siri. Their purchases included:

1) a mix of LIFX and Phillips Hue light bulbs for remote-control lighting;
2) a TP-Link power switch to control their appliances; and
3) A Hello Barbie talking doll.

In the fourth scenario, a trendy young city couple place a high priority on their social life. Eddie and Jenny like to listen to music in every room of their home, including on their rooftop terrace. They also spend a lot of time on their mobile devices, and subscribe to the major movie-streaming services. Jenny likes watching the latest movies while Eddie prefers playing computer games. Both have busy professional lives and often work nights and on weekends. They have bought the following devices:

1) Smart TV with Google Chromecast, which plays games and streams videos;
2) Triby portable speaker;
3) Amazon Echo voice-activated assistant;
4) HP Envy smart printer; and
5) Pixstar photo frame, which automatically syncs photos with their Facebook accounts.

Testing

We selected a number of devices based on the above scenarios as well as on product availability and popularity in Australia, and carried out detailed tests on each (as well as its supplied mobile app and data server). These tests ranged from the simple (capturing wireless transmissions from the device to evaluating the contents of the communication) to the complex (making the device communicate to a fake server, and overwhelming the device with fake query messages). We automated the process in a laboratory to make it easier to reproduce and compare results.

The IoT devices were connected to a home gateway router either through Wi-Fi or via direct connection with an Ethernet cable. The applications for the IoT devices were downloaded onto an Android tablet, which was connected to the same router. Checks were performed from a laptop running a digital testing platform called Kali Linux, which was on the same network as the IoT devices.

Using this setup, we ran basic computerized scripts and penetration testing tools to assess the safety and security performance of each IoT device.

The devices tested were:

- Cameras (TP-Link, Belkin, Dlink, Samsung, Canary, Netatmo and Nest Drop).
- Motion sensor (Belkin).
- Smoke alarm (Nest).
- Medical device (Withings sleep monitor, Withings weighing scale).
- Air quality monitor (Awair, Netatmo weather station).
- Light bulbs (Phillips Hue and LIFX).
- Power switches (Belkin and TP-Link).
- Talking doll (Hello Barbie).
- Photo frame (Pixstar).
- Printer (HP Envy).
- Controller (Samsung SmartThings).
- Voice assistant (Amazon Echo).
- Smart TV with Google Chromecast.
- Speaker (Triby portable speaker).

The Results section lists full tables of results showing how each device performed in each category. The results of our tests were consistent and alarming. Every device we tested showed some form of vulnerability in integrity, access control, or reflection capabilities. Many were susceptible to attack in a number of ways. The Phillips Hue light bulb and Belkin switch had notably poor security. But there was some good news. Devices such as the Amazon Echo, Hello Barbie, Nest Drop Cam, and Withings sleep monitor were relatively secure in terms of confidentiality. The Echo, in particular, was a top-rated device in security with encrypted communication channels and almost all of its ports closed to outside attack. A vivid illustration of these vulnerabilities can be gained by applying them to our four scenarios.

In the first scenario, a former target of Tuan's investigation would be able to sit in a car outside her house and deduce her Wi-Fi network password using freely available software. He would then place a cheap battery-powered device beneath her letterbox. This device connects with her home wireless network, capturing all of the information being transmitted by her IoT devices. This information is then sent back to his laptop, which he monitors from his home. Essentially, his device is performing a "man-in-the-middle" attack on Tuan's motion sensor and camera — both of which send out information that is not encrypted. This makes it quite simple to see video and read motion-sensor information from Tuan's devices on his laptop at home. He would therefore know when Tuan's devices have been inactive for a few hours. Surmising that Tuan is away, perhaps in Melbourne or Sydney, he drives back to his parking spot in the street outside Tuan's home. He uses a denial-of-service attack on Tuan's motion sensor, cameras, and smoke alarm by bombarding them with a large number of requests. Unable to cope, these devices simply shut down. This ensures that she will never get the smoke alert from her IoT alarm — even though her home has been physically set alight.

In the second scenario, a criminal buys a list of email addresses of people who have recently registered IoT products. One of these belongs to Joe and Lorna Jones. The criminal sends them an email that contains a link to an app that promises technology customers help with their finances. The app, however, has embedded malware that scouts for IoT devices. Lorna is not sure what the email is about but thinks it sounds interesting. Without thinking, she manages to download the app. The malware immediately disables the Joneses' firewall and enables port forwarding, making them vulnerable to security breaches. Now the criminal is in control. His malware finds unencrypted messages from their weighing scales, enabling him to deduce their names, ages, gender, height and weight. From this, he can start hatching a plan for someone else in his criminal syndicate to steal the Joneses' identity and take their social security benefits. He can also use Joe and Lorna's IoT devices to reflect and amplify attacks on other Internet-connected devices. Whenever he likes, he can use the open ports on the Joneses' Withings sleep monitor, Awair air quality monitor, and Netatmo weather station and use them as part of a network of compromised devices to launch massive cyber-attacks. Note, however, that in general, health monitoring IoT devices do not tend to have many security problems. Although the Awair air quality monitor could stop functioning if it's forced to deal with a large amount of Internet traffic, it encrypts all data sent to the server.

In the third scenario, an opportunistic neighbor sees the Singhs as a potential soft burglary target. He uses a remote device to deliver malware that snoops on local Wi-Fi traffic. The Singhs' IoT devices, especially their power switch and lights, provide a good indication of their presence in, or absence from, their home. More importantly, the neighbor can alter the state of the devices. The Phillips Hue light bulbs do not send encrypted information, so he can turn them on or off and change their color and brightness. The LIFX bulbs have encrypted messages but they can be decrypted with little effort. The TP-Link power switch also uses encrypted data but has a very weak key; it can be broken easily. Under certain conditions, the Hello Barbie doll enables outsiders to listen in on conversations while the doll's talk button is pushed.

In the fourth scenario, a cyber-stalker uses a password-cracking tool to gain access to Eddie and Jenny's Wi-Fi network. Like many others, they have not changed the default username or password ("admin") on most of their devices. Once in, the stalker can use simple request functions to get information on what videos and games they play through Google Chromecast — she might even be able to post a threatening text or video on their television screen. She knows their printer is particularly vulnerable. Using the basic Internet Printing Protocol, she can see any documents they have scanned recently or might even print a threatening or obscene message on the device. Although most of Eddie and Jenny's devices are relatively safe compared with other IoTs tested, the HP Envy printer is an exception. It has poor security protection, with many open ports that are not protected by a password, allowing an attacker easy access. It also allows an attacker to print documents or stop others from printing entirely.

Evaluate

We invited IoT suppliers, consumers, insurers, and regulators to evaluate our results at a workshop. In this section, we discuss their reactions and expectations.

A frequent theme among attendees was that consumer expectations must survive a transition to the digital age. Most consumers of smart-home IoT devices will not scrutinize manufacturers' license agreements, and they cannot be expected to as the agreements are frequently complex and unlikely to be enforced. They assume that manufacturers or service providers will supply any software updates necessary to continue running their applications. Similarly, consumers expect that a smart-home device placed on their home network will not create a backdoor to other devices in their home. More generally, they expect that technical security is someone else's responsibility.

We believe this expectation is reasonable in light of consumers' experiences with non-IoT products. Car buyers, for instance, are only required to ensure that their cars are locked, perhaps parked in a secure garage, and regularly serviced in line with the manufacturer's specifications. They are not expected to also be automotive engineers, mechanics or locksmiths. And yet, the question persists: how much education is required for a consumer to know that their IoT devices are "safe"? It's possible to foresee the use of a security "star rating" for IoT devices — similar to energy- or water-efficiency ratings on household appliances — that may allow consumers to make informed purchasing decisions. Such a ratings scheme might enable market forces to decide how important the security and safety of IoT devices are to consumers [8].

Such a scheme is not without complexity of its own. Security ratings, after all, cannot be static, since security threats evolve continuously. The implications of a low security star rating may be unclear to consumers.

Further, the issue of data ownership and its sharing remains murky [9]. Consumers may expect their service providers will not on-sell data generated by their smart-home IoT devices, for example, despite some license agreements allowing just that. Any ratings system, and improvements to consumer decision making, need to take this into account.

For manufacturers, a major gap exists between consumers' expectations that IoT devices will be kept up-to-date with near-invisible software "patching" and the current reality that many devices simply cannot be updated. While smartphones can be patched with regular updates, the firmware in many IoT devices cannot be patched due to small memory capacity, lack of a management system, the transient nature of network connectivity, or some other issue. In the cases where devices can be updated, the technical demands required to make this happen are beyond the ability of most consumers.

Furthermore, in a world of disarticulated production, it is simply not clear who is most responsible for a security shortfall: is it the company that designs the device, or the one that supplies component software? Or is it the company that supplies the network in which the device is embedded?

Further, manufacturers often focus on price competitiveness rather than security, especially because development costs in this area are high. They are more likely to move quickly to the next, more advanced version of their models because that is where the greatest profit lies. The performance of previous models is not likely to concern them, particularly once they're out of warranty. Manufacturers are also aware that consumers who own webcams and digital video recorders used in DDoS attacks do not personally know the victims, and are not likely to pay too much attention to security features. In such cases, security is something that affects people who are not involved in the transaction between buyer and seller — an "externality" in economic terms.

Insurers should reconsider their approach to manufacturers and consumers of IoT devices. The cyber insurance market is said to be worth $3 billion to $4 billion per year, and is growing at 60 percent annually [10]. Companies that sell IoT devices may need to be insured against the possibility that their products may cause harm to their customers, or others. Effective policy is needed to ensure businesses that produce devices unfit for purpose, or that are repeatedly hacked, cannot continue to do so. A business that is compromised, but has taken reasonable steps to resolve the issue — and shows no negligence — should be able to claim on its insurance.

Recently IoT devices have also been made available for extremely intimate and sexual applications with devices enabling remote logging and control [11], even incorporating cameras. In this context other security researchers have identified significant flaws in the implementation of connectivity, privacy, and data management, which they argue is through the poor choice of source code reused from public repositories [12]. In one case privacy protections in the U.S. meant that customers could receive compensation for breaches of their usage data after a court finding that the breach had not been disclosed to customers.

In this context the potential for serious sexual assault leaves device manufacturers clearly open to adverse judgement and reputational damage even if perpetrators of such crimes are difficult to identify and pursue.

For these and other reasons, there may be no feasible market based solution to the issue of poor IoT security, meaning the onus may fall on regulators.

Proposal

Resolution of the security risks identified in our study is hampered by the siloed nature of regulation that is now becoming more broadly applicable due to the expansion of communications and forming the IoT. Functions and objects are the responsibility of discrete government departments and regulatory agencies, but the agencies now find themselves potentially responsible for new areas. Further exacerbating this problem is that regulatory standards and benchmarks that apply in one jurisdiction do not necessarily apply within another.

Medical, traffic control, and building management systems, cameras, light bulbs and cars with driver-assist features use an increasing number of IoT devices, yet are regulated by separate government departments. In Australia for example, the Therapeutic Goods Administration within the Department of Health regulates medical devices, whereas the Australian Communications and Media Authority regulates telecommunications, broadcasting, radio communications, and the Internet, and the Australian Competition and Consumer Commission regulates consumer safety and fair trade. Regulating IoT devices will involve input from elements within each of these entities, and complexity is only likely to increase over time. The Australian government Department of Infrastructure and Regional Development regulates vehicle safety, and may require real-time access to data feeds from vehicles using IoT devices. As driver-assistance technologies develop in cars, the need for cross-departmental attention will increase. As in Australia, today's regulatory agencies across the world were created to respond to the rise of earlier technologies. The coming IoT revolution will require new regulatory expertise that cuts across the current set of agencies.

We therefore propose a more coordinated and exhortative approach to regulation. Manufacturers will need to be encouraged to build security at the design phase. A "security by default" attitude would see consumers having to deliberately disable rather than deliberately enable security features. A mechanism may need to be found to coordinate software updates among third-party vendors, and to facilitate the coordinated disclosure of vulnerabilities. Here, a role may be found for national cybersecurity agencies, such as the Australian Cyber Security Centre, to coordinate the security knowledge-sharing of developers, manufacturers, and service providers.

Bodies and services that may have been exempt in the past from regulation may also come under future scrutiny due to the evolving need for consumer and community protection. Because of the serious threat to infrastructure, it is conceivable that governments may in the future require Internet service provider networks to comply with network security standards or meet performance benchmarks. Devices provided by manufacturers or Internet service providers to perform network boundary roles, such as home gateways, could be expected to come under higher levels of requirements. This would mean devices shipped with default passwords, for example, could become a thing of the past.

Further research along the lines of the STEP model is needed in order to continue to shed light on the burgeoning field of IoT devices.

Results

Based on the major threats we identified, Figures 1-4 show how each IoT device performed in the four categories — confidentiality, integrity and authentication, access control, and the ability to withstand reflective attacks.

From this, we gave each device an overall rating for each category. If a device passed a test it was rated "good" (represented by green "A" boxes in the tables); if it failed it was "poor" (red "C" boxes). If it did not pass the test but the attack was unsuccessful, it was rated as average (yellow "B" boxes). The grey boxes

Confidentiality										
	Device to Server			Device to Allocation			Application to Device			All
Devices	Plain Text	Protocol	Entropy	Plain Text	Protocol	Entropy	Plain Text	Protocol	Entropy	Privacy
Phillip Hue Light Bulb	A	A	A	C	C	C	A	A	A	C
Belkin Switch	B	▨	A	C	C	C	A	A	A	C
Samsung Smart Cam	A	▨	A	A	A	A	A	A	A	A
Belkin Smart Cam	A	▨	A	A	A	A	A	A	A	A
Awair Air Monitor	A	A	A	A	A	A	A	A	A	A
HP Envy Printer	A	A	A	C	C	C	A	A	A	C
LIFX Bulb	A	A	A	A	▨	C	A	A	A	A
Canary Camera	A	A	A	A	A	A	A	A	A	A
TP Link Switch	A	▨	A	A	▨	C	A	A	A	A
Amazon Echo	A	A	A	A	A	A	A	A	A	A
Samsung Smart Things	A	A	A	A	A	A	A	A	A	A
Pixstar Photo Frame	A	A	A	A	A	A	A	A	A	A
TP Link Camera	A	▨	A	C	C	A	A	A	A	C
Belkin Motion Sensor	A	A	A	C	C	C	A	A	A	C
Nest Smoke Alarm	A	▨	A	A	A	A	A	A	A	A
Netatmo Camera	A	A	A	B	C	A	A	A	A	A
Dlink Camera	C	C	C	A	A	A	A	A	A	A
Hello Barbie Companion	A	A	A	A	A	A	A	A	A	A
Withings Sleep Monitor	A	▨	A	A	A	A	A	A	A	A
Nest Drop Camera	A	A	A	A	A	A	A	A	A	A
Netatmo Weather Station	A	A	A	A	A	A	▨	▨	▨	A
Triby Speaker	A	A	A	A	A	A	A	A	A	A
Withings Weighing Scale	C	C	C	A	A	A	C	C	C	C
Chromecast	A	A	A	C	C	C	A	A	A	C

FIGURE 1. Confidentiality rating.

show when a particular attribute could not be tested or assessed.

Note these tests were performed at a point in time and may have been improved or further deteriorated since the date of testing in April 2017.

Confidentiality Rating

Confidentially is a measure of the security of data running between the IoT device, the router, and our server.

Our tests show whether the communications sent and received were encrypted (the most difficult to read), encoded (hard but not impossible), or plain text (easiest to hack).

Figure 1 shows how each device performed in confidentiality testing.

- Most of the devices had fairly secure communications in two channels (device to server and user app to server) but were vulnerable when they communicated with their user app.
- Five of the devices — the Phillips Hue light bulb, Belkin switch and motion sensor, HP Envy printer, and TP-Link camera — sent data in plain text rather than encrypted code. This would make it relatively simple for hackers to deduce when a user is at home, based on whether the power switch is on or off, or when the light bulb was last used, for example.
- The TP-Link camera was particularly susceptible to attack. Not only might an attacker view any video and audio footage based on reassembled data, the default authentication password "admin" was easily decoded.

Integrity Rating

We checked the integrity and authentication of each device by setting up a fake server to "listen" on the port used by the real server. This technique is known as a "man in the middle attack."

Using a number of methods, this fake server communicated with each device to see if it could be authenticated. We also tested to see if the devices could be controlled by outside influences.

Figure 2 shows how each device performed in integrity testing.

- These results show that all of the IoT devices were vulnerable to an attack through the Domain Name System (DNS) protocol. This means that attackers could hijack the system and impersonate the legitimate server of the IoT device. They would be protected, however, through proper authentication.
- The two light bulbs that were tested communicated with the fake server, which is a concern.

Access Control Rating

We tested to see if any ports on a device were "open," allowing the port to be exploited by attackers. Based on this, we launched a password-guessing attack to see if they were protected by strong security protocols.

Each device was also checked to see how much traffic any open ports could handle before they were brought down in a DDoS attack.

Integrity and Authentication				
Devices	Replay Attack	DNSSEC	DNS Spoofing	Fake Server
Phillips Hue Light Bulb	C	C	C	C
Belkin Switch	C	C	C	C
Samsung Smart Cam	A	C	C	A
Belkin Smart Cam	A	C	C	A
Awair Air Monitor	A	C	C	A
HP Envy Printer	C	C	C	A
LIFX Bulb	C	C	C	C
Canary Camera	A	C	C	A
TP-Link Switch	C	C	C	A
Amazon Echo	A	C	C	A
Samsung Smart Things	A	C	C	A
Pixstar Photo Frame	A	C	C	A
TP Link Camera	A	C	C	A
Belkin Motion Sensor	A			
Nest Smoke Alarm	A	C	C	A
Netatmo Camera	A	C	C	A
Dlink Camera	A			
Hello Barbie Companion	A	C	C	A
Withings Sleep Monitor	A	C	C	A
Nest Drop Camera	A	C	C	A
Netatmo Weather Station		C	C	A
Triby Speaker	A	C	C	
Withings Weighing Scale		C	C	
Chromecast	C	C	C	A

Key:
DNS: Domain Name System
DNSSEC: DNS Security Extensions

FIGURE 2. Integrity and authentication.

Access Control							
Devices	Open Ports (TCP)	Open Ports (UDP)	Vulnerable Ports	Weak Passwords	ICMP DDoS	UDP DDoS	Num. of TCP Connections
Phillips Hue Light Bulb	C	C	C	A	B	C	C
Belkin Switch	C	C	A	A	C	C	C
Samsung Smart Cam	C	C	C	A	C	C	C
Belkin Smart Cam	C	C	C	A	C	B	C
Awair Air Monitor	B	B	A	A	C	C	A
HP Envy Printer	C	C	C	A	A	A	C
LIFX Bulb	A	B	A	A	C	B	A
Canary Camera	A	A	A	A	C	A	A
TP-Link Switch	C	C	C	A	C	C	C
Amazon Echo	C	C	A	A	B	C	C
Samsung Smart Things	C	B	C	A	C	C	C
Pixstar Photo Frame	A	C	A	A			A
TP Link Camera	C	C	C	C	C	B	C
Belkin Motion Sensor	C	C	A	A	C	B	C
Nest Smoke Alarm	B	C	A	A			A
Netatmo Camera	C	C	C	A	C	B	C
Dlink Camera	C	C	C	C	C	B	C
Hello Barbie Companion	C	A	A	A	C	A	A
Withings Sleep Monitor	C	C	C	A			C
Nest Drop Camera	A	B	A	A	C	A	A
Netatmo Weather Station			A	A			
Triby Speaker	C		A	A	C		C
Withings Weighing Scale	A		A	A	A	A	A
Chromecast	A		A	A	C		C

Key:
TCP: Transmission Control Protocol
UDP: User Datagram Protocol
ICMP: Internet Control Message Protocol
DDoS: Dedicated Denial of Service

FIGURE 3. Access control.

Figure 3 below shows how each device performed in the access control testing.

- Almost all of the devices had some form of open-port vulnerability. This would enable intruders to communicate with or gain access to the devices.
- Both the Belkin Smart Cam and HP Envy printer exposed a wide range of open ports.
- Disturbingly, both the HP printer and DLink camera had no protection for remote access.
- The last three columns show that most of the devices were susceptible to at least one form of DDoS attack.

Reflection Attack Rating

We evaluated all of the devices in their ability to "reflect" traffic and overload a victim's network, forcing it to shut down.

"Amplification" is a type of reflection attack [13]. In this case, the reflection is achieved by gaining a response from an innocent IoT device to a spoofed IP address (a victim machine or server). During an amplification attack, an attacker sends a query with a forged IP address (the victim's) to the reflector (the IoT device), prompting it to reply to that address with a response. With numerous fake queries being sent out, and with several IoT devices replying simultaneously, the victim's network is overwhelmed by the sheer number of responses it's asked to make.

Figure 4 below shows how each device performed.

- Most of the devices were unable to withstand an ICMP reflection attack.
- All devices, except the LIFX light bulb, were susceptible to reflecting some form of attack.
- The Samsung Smart Cam was vulnerable across a number of protocols.

Reflection Attacks				
Devices	ICMP Reflection	SSDP Reflection	SNMP Reflection	SNMP Public Community String
Phillips Hue Light Bulb	C	C	A	A
Belkin Switch	C	C	A	A
Samsung Smart Cam	C	A	C	C
Belkin Smart Cam	C	C	A	A
Awair Air Monitor	C	A	A	A
HP Envy Printer	C	A	C	A
LIFX Bulb	A	A	A	A
Canary Camera	C	A	A	A
TP Link Switch	C	A	A	A
Amazon Echo	C	A	A	A
Samsung Smart Things	C	A	A	A
Pixstar Photo Frame	C	A	A	A
TP-Link Camera	C	A	A	A
Belkin Motion Sensor	C	C	A	A
Nest Smoke Alarm	C	A	A	A
Netatmo Camera	C	A	A	A
Dlink Camera	C	C	A	A
Hello Barbie Companion	C	A	A	A
Withings Sleep Monitor	C	A	A	A
Nest Drop Camera	C	A	A	A
Netatmo Weather Station	▨	A	A	A
Triby Speaker	C	A	A	A
Withings Weighing Scale	🟩	A	A	A
Chromecast	C	A	A	A

Key:
ICMP: Internet Control Message Protocol
SSDP: Simple Service Discovery Protocol
SNMP: Simple Network Management Protocol

FIGURE 4. Reflection attack.

Current Generation of IoT Devices Vulnerable to Attack

Consumer products connected to the Internet will soon become commonplace in homes and businesses, and will offer customers many productivity and lifestyle benefits. Our study, however, suggests that the current generation of IoT devices is vulnerable to attack in a number of ways. It is a complex problem, and there don't appear to be any "single bullet" solutions to make IoT devices safer or more secure. We hope this article sets the platform for a dialogue between consumers, suppliers, regulators, and insurers of IoT devices to develop appropriate methods to tackle the problem.

Author Information

Vijay Sivaraman and *Hassan Habibi Gharakheili* are with the School of Electrical Engineering and Telecommunications, University of New South Wales (UNSW), Australia.

Clinton Fernandes is with the School of Humanities and Social Sciences at UNSW and the Australian Centre for Cyber Security, Australia.

Narelle Clark and *Tanya Karliychuk* are with the Australian Communications Consumer Action Network, Australia. Email: research@accan.org.au.

References

[1] D. Evans, "The Internet of Things: How the next evolution of the Internet is changing everything," CISCO,White Paper, 2011; https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.

[2] J. Chambers, Executive Director of Product Innovation, comments presented at UNSW workshop (Australia), Apr. 20, 2017.

[3] N. Dhanjani, *Abusing the Internet of Things: Blackouts, Freakouts, and Stakeouts*. O'Reilly Media, 2015.

[4] E. Fernandes, J. Jung, and A. Prakash, "Security analysis of emerging smart home applications," in *Proc. IEEE Symp. Security and Privacy* (San Jose, CA, USA), May 2016.

[5] F. Loi, A. Sivanathan, H. Habibi Gharakheili, A. Radford, and V. Sivaraman, "Systematically evaluating security and privacy for consumer IoT devices," in *Proc. ACM CCS Workshop IoT Security and Privacy* (Texas, U.S.A.), Nov. 2017.

[6] Parliamentary Joint Committee on Intelligence and Security, *Advisory report on the Telecommunications (Interception and Access) Amendment (Data Retention) Bill 2014*. Canberra, Australia: Commonwealth Parliament, 2015, p. 11.

[7] C. Fernandes and V. Sivaraman, "It's only the beginning: Metadata retention laws and the Internet of Things," *Australian J. Telecommunications and the Digital Economy*, vol. 3, no. 3, Sept. 2015.

[8] ZDNet, "No stars for Internet of Things security," presented at AusCERT 2016 Conf., May 27, 2016.

[9] "The data economy: Fuel of the future," *The Economist*, May 6, 2017.

[10] "The myth of cyber-security" & "Why everything is hackable," *The Economist*, Apr. 8, 2017.

[11] M. Wynn et al., "How to practice safe IoT: Sexual intimacy in the age of smart devices," in *Proc. ACM CCS Workshop on IoT Security and Privacy* (Texas, USA), Nov. 2017.

[12] R. Chirgwin, "Wi-Fi sex toy with built-in camera fails penetration test," *The Register*, Apr. 4, 2014; https://www.theregister.co.uk/2017/04/04/intimate_adult_toy_fails_penetration_test/.

[13] M. Lyu et al., "Quantifying the reflective DDoS attack capability," in *Proc. ACM* (Boston, MA, U.S.A.), Jul. 2017.

Algorithmic Governance in Smart Cities

The Conundrum and the Potential of Pervasive Computing Solutions

Franco Zambonelli, Flora Salim,
Seng W. Loke, Wolfgang De Meuter,
and Salil Kanhere

Pervasive and mobile computing technologies can make our everyday living environments and our cities "smart", i.e., capable of reaching awareness of physical and social processes and of dynamically affecting them in a purposeful way [1]. This is already happening, e.g., in the form of digital traffic signs that suggest in real-time the best traffic directions or the availability of parking spots, and also in the form of location-based social networks that inform us about noteworthy events. Soon we expect that the pervasive diffusion of sensing, actuation, and computing will allow our urban environment to fully self-regulate in autonomy most of its processes, and to guide and support our everyday activities.

It is generally acknowledged that living in a smart environment makes us smarter by increasing our overall levels of awareness of ongoing urban activities [2]. Also, by supporting and facilitating our customary activities (e.g., driving, finding information, and goods), living in a smart environment can make life much more pleasant and less stressful, and also make the environment more sustainable.

MNBB/ISTOCK

However, the evolution of smart environments also carries potential risks for individuals and for society as a whole. In particular, if most of our everyday activities can be automated, we could be tempted to increasingly delegate the governance of such activities, and the governance of the whole city, to the algorithmic engines of the smart city infrastructure. Thus, rather than taking advantage of the augmented capabilities of perception and participation enabled by the technologies, we could end up losing critical attention, abandoning individual decision making for relying on collective computational governance of our activity, losing awareness of environmental and social processes, and ultimately lose power and become dumb [3].

Digital Object Identifier 10.1109/MTS.2018.2826080
Date of publication: 31 May 2018

In this article we elaborate on the key concepts of algorithmic governance in smart cities, discussing its likely increasing role in the future. Without any intention of dramatizing or of embracing dystopian visions, we intend to outline some specific problems that could potentially occur in future smart cities, and eventually analyze some key directions to prevent or mitigate — also with the fundamental support of pervasive computing technologies — these potential problems and possibly turn them into advantages.

Smart Cities: From Citizen Support to Algorithmic Governance

Algorithmic governance, in general terms, concerns empowering software to take decisions and to autonomously — i.e., without human supervision — regulate some aspects of our everyday human activities or some aspect of the society, according to some algorithmically defined policies [4]–[6].

We are already subject to algorithmic governance in a variety of different aspects of our personal and social lives. Google search dictates what information we find on the Internet, as we typically accept the suggestions appearing on the first page. Facebook news feed algorithms dictates what are the relevant posts to show us, and given that young people rely on Facebook as the primary source of news, this implies that they have fully delegated to Facebook the activity of seeking and filtering information [7].

Moving from the individual to the societal sphere, examples of algorithmic governance can be found in trading, where most of decisions (and thus the oscillations of markets and our finances) now rely on complex agent-based decision making; or in the pricing strategies of airlines that rely on complex analysis of travel trends.

In the area or urban management, some early examples of algorithmic governance can be found in traffic management (e.g., traffic lights that adaptively their frequency depending on the sensed traffic flow), public transport (e.g., to adapt bus schedule and routes to meet the transport demand in real time), and energy management (e.g., to automatically tune energy pricing depending on the instantaneous balance between supply and demand).

Current examples of algorithmic governance for smart cities still rely on rather limited capabilities of sensing and actuating. However, the increasing spread of pervasive computing and Internet of things (IoT) technologies will soon make it possible to sense at incredible levels of detail every single event happening in every corner of a city, and to trigger flexible actions to affect the state of things via a variety of actuators, robots, and autonomous vehicles [1]. The assessed mid-term future for urban mobility is the one in which citizens and merchandise will be carried to any desired destination via myriads of self-driving vehicles, globally orchestrating their movements and routes with each other and with the urban street infrastructure. Similarly, the flow of pedestrians will be somehow steered (via digital signages or apps on wearable devices) and orchestrated so as to avoid dangerous situations [8].

Pushing the vision forward, we can imagine that the entire management of cities will be soon governed and actuated in an automatic, unsupervised way. For instance, this can include the management of waste collection (which also includes the possibility to make citizens individually accountable for what they produce as waste), decisions on new urbanization, and management of roads and other infrastructure (e.g., by automatically deciding which roads and traffic lights to replace depending on their actual state and available budget).

In general, living in a smart environment and being made part of its awareness can potentially make us smarter. That is, it can notably increase our perception and social capabilities [2], and our ability to understand situations and react to them. In addition, allowing our everyday urban lives to be governed by some automated software systems promises to notably increase our quality of life. In fact, it will relieve us from a number of boring physical and mental activities, and enable us to do much more interesting things that, say, driving and having to decide on a route to a destination. In other words, it will make it possible to satisfy needs at the highest levels of the Maslow pyramid [9]. However, as we will elaborate in the following, algorithmic governance also comes with a number of potential dangers, and raises the risk of actually making us "dumber" rather than smarter.

Possible Perils of Algorithmic Governance

The primary source of all peril related to algorithmic governance is that, very often, software and algorithms are designed as "black boxes" with little understanding of how they actually work. This lack of understanding not only involves the final users (i.e., in case of smart cities, all citizens) but quite often also the stakeholders (e.g., municipalities and decision makers), and the developers, if they do not care enough to understand how these algorithms operate and have a discussion with designers and stakeholders.

This limited understanding of algorithms, from what we can see so far, does not prevent people from relying on them for various activities. For instance, people fully rely on Google's search results, even without knowing anything about its underlying PageRank algorithms. This is the key difference between algorithms and other classes of technologies: we are using them and relying on their judgement and suggestions without knowing the reasons we are being suggested something. In Smart Cities, a pervasive environment governed by algorithms will make us apparently smarter in our capabilities, but to some extent will also make us "dumber" in that our actions will no longer be conscious. For instance, roaming in an unknown city with a paper map enables us to absorb the basic fabric of the city and, on this basis, to consciously decide what route to take. Conversely, roaming with a GPS navigator does not require knowing anything about the city structure, and we are willing to accept route suggestions without investigating further the reason a certain route has been selected. In other words, our notion of "dumber" here is in the sense of potentially losing the ability to make good judgements in some situations, reducing discernment, and lazily deferring to the algorithm. The overreliance on algorithms, which are capable (unlike other technologies) of making judgements or decisions on our behalf, without our questioning or thinking them through, will make us increasingly depend on the algorithms in our everyday life, instead of depending on human rationale and our own reasoning.

One might say that the above issue will not cause any trouble because algorithms will be developed to serve citizens, for their own good and in accord with rules and policies of the municipality, and that such rules will be made transparent. Yet, as we elaborate in the following: 1) decisions made by algorithms may be biased and have inaccurate information; 2) there is risk that someone manage to make intentional misuse of algorithms without no one noticing; 3) politics and decision makers themselves may end up having little clue as to the actual working of algorithms, thus losing power.

With regard to the first issue, we normally trust the developers of the technologies we use, knowing that societal norms, regulations, and reputation largely help to protect that trust. However, intentional or (more often) unintentional bias, errors, or incomplete information, can subtly hide in algorithms, along with biases in behavior due to values (right or wrong) that developers or creators of the technology might have. This issue could be exacerbated by the increasing influence and role of technology in daily decision-making. For example, bias can creep into a machine learning algorithm [10] due to wrong or inadequate datasets, rather than any malice on the part of the developers (e.g., a system that is racially biased against people with black-sounding names when judging potential for crime, or software that could not do face recognition well for people with dark skin, or unfair judgments on insurance or loans, all due to issues with data used in training the algorithms). Also, what is a "normal person" or "normal behavior" is often contextual and hard to define — which can influence default settings in software [11], e.g., gender bias in default character profiles in some games, or someone struggling to read a critical alert message in a non-native language, or what an algorithm portrays as "normal looking."

Algorithms also could have incomplete information in decision-making or lack knowledge of exceptional cases even if developers do endeavor to be comprehensive. In some cases, even if tempted to doubt the algorithm, profitability in following the algorithm's decision can be high, but not necessarily by virtue of how good the algorithm is. For instance, with smart algorithms performing sentencing, or the "robotization of justice" [12], by deciding against an algorithm's position for guilt, one personally takes on the responsibility for a possible second offense. In any case, even if the thought that algorithms can do better than humans is becoming prevalent, and despite the recent breakthroughs in deep learning, for neural networks to achieve the computational level of a human brain, neural networks still need far more power [13].

With regard to the second issue, independently of the reliability and trustability of algorithms, there is an issue of who is devoted to managing such algorithms and of how we can trust those people to manage them. For instance, developers could be influenced by imperatives from (e.g., an abusive) government, when developing technology. If there is a problem with the current government, algorithmic technology could be intentionally misused, hiding behind the "veil of sophistication." In particular, if algorithms are left to their own devices to govern Smart Cities, the more complex conundrum is what algorithmic government models for Smart Cities should look like to prevent or mitigate voluntary or unintended abuses. Who will be in charge of deciding which algorithms to run? If politics are in charge of deciding the rules governing a city, who will be in charge of developing algorithms and to check that they are adhering to the rules? Who will tune the parameters for the algorithms and how? What about the possibility of instantiating multiple algorithms, and that such algorithms become antagonistic [14], i.e., conflicting with each other? What does an objective function for governance models look like? What does a termination condition for algorithmic governance models look like? Who will be responsible for bugs? For algorithms governing safety critical situations [31], such questions are already pressing, and this problem will be dramatically exacerbated in the future. Just think at the use of artificial intelligence (AI) and autonomous decisions in weaponry [15]. Hinton, in [13], also argued that there is not only a need to tame AI research, but also to improve political systems so that AI is not misused.

Strictly related, algorithms governing our cities may also be involved in decisions involving ethical or moral dilemmas. For instance, a human driver who managed to swerve her car in time to avoid killing several pedestrians while sacrificing herself might be lauded, but a self-driving car that killed its driver even while saving pedestrians might worry passengers [16]. Self-driving cars could reduce the need to learn driving, and so fewer people might end up knowing how to drive. This could be a problem in situations where such a skill is indeed needed. But moral algorithms that make human-accepted judgments are problematic. For example, an algorithm that behaves in a utilitarian manner could benefit society as a whole, but could hurt individuals, and therefore might not be accepted.

With regard to the third issue, the risk that politicians and governments lose control over algorithms, can lead to an "algocracy." The term "algocracy," contrasts with "democracy," and literally means that the power ("kratos" in greece) lies with the algorithms rather than with the people ("demos" in Greece). To some extent, it may appear we are already living in a partial algocracy, given that algorithmic decisions already affect some of our civil life. For instance, when applying for a visa for some countries, based on some business rules, applicants with a certain type of passport will receive immediate clearance, whereas others will be pushed to the next level of scrutiny. These procedures treat people differently based on the settings and semantics of algorithms. Yet until the settings and the semantic of such algorithms are perfectly compliant with legislation and rules, the power still resides with the government and, ultimately, in a democratic system, with the people who voted for them.

However, given that algorithms can be difficult to implement, configure, and fully understand, the risk exists that governments end up relying on algorithms they do not fully understand without being capable of effectively verifying the adherence of the algorithms to the existing laws. Thus, we may end up implicitly delegating decisional power to the algorithms, or to the group of people devoted to designing and developing them.

Algorithms in future societies and cities will serve the same role that civil law and urban regulations, respectively, serve in today's democratic systems. Accordingly, new political procedures need to be put in place to regulate which code is installed. The responsibility for (technical) code verification and (juridical) semantic verification needs to be clearly defined. Most importantly, to avoid having algorithmic governance degenerate into algocracy, it is necessary for citizens, politicians, and decision makers to become capable of understanding and harnessing the complexity of algorithms and their configuration.

How to Deal with the Problems

Let us now analyze what solutions, possibly enabled by the same pervasive computing technologies that cause them, can be envisioned to attack the identified problems.

Data Access Control

Algorithmic governance in Smart Cities is enabled by the availability of large amounts of data, making it possible for algorithms to understand what is happening and act accordingly. In this context, the dense and pervasive collection, processing, and dissemination of data in the midst of people's private lives, while useful for offering a range of sophisticated and personalized services that provide utility to users, necessarily gives rise to certain privacy and algorithmic concerns.

From the privacy viewpoint, the pervasive collection of information exacerbates existing issues associated with privacy in data handling. In fact, such details can be used to algorithmically construct a virtual biography of our activities, revealing private behavior and lifestyle patterns. Disclosure of this type of analysis or information gives rise to the notion of behavioral privacy [17] which is distinctly different from traditional identity privacy. The possession of such detailed personal information about an individual may confer power over that individual, resulting in potential misuse by governments, corporations, or other individuals. For example, households are being equipped with smart meters to act as providers of temporally detailed energy consumption reports. The utility companies use the data to better estimate domestic power consumption leading to optimized distribution. However, as shown in [18], several unintended and sensitive inferences such as occupancy and lifestyle patterns of the occupants can be made from the data. Accordingly, we need configurable privacy-preserving tools that afford users fine-grained control over how their personal information is shared.

From the algorithmic viewpoint, it is of fundamental importance for users to understand the type of personal information being used by algorithms to make decisions, and how such information is used. In addition, in cases where the user understands that such information is not correct, is biased, or is not appropriately used (and thus leads to incorrect algorithmic behaviors), the user can exploit the above-mentioned configurable privacy-preserving tools to adjust the usage of information by algorithms. For instance, consider an automated home heating system that self-regulates based on the life patterns of inhabitants (and bills accordingly), and a person who was constrained at home for 15 days due to a bad winter flu. When such person eventually goes back to work, he should be able to "see" if the heating system is still acting on the basis of the wrong assumption that he is at home with a flu. The risk, otherwise, is to lazily (and stupidly) accept paying more for heating.

For both concerns, pervasive computing technologies can potentially enable users to access the appropriate sensors (e.g., wirelessly via their mobile phones), and see what data they have produced, and how they have used what algorithms.

Algorithmic Guardians

If today algorithms affect what we can view online, tomorrow they will modify our physical reality at home or in a Smart City, and — as stated earlier — will do that in personalized way. However, even if we are offered the possibility of seeing what data is being used and by what algorithms, the resulting personalization process might not be transparent or comprehensible, and it could be hard to understand and interpret, especially for non-data-literate users.

Besides the problem of understanding personalization, another issue that may arise concerns the fact that personalization might not always serve the users' interest, but rather the interests of the algorithms' creator. There is usually an interest gap between you, theuser, and the third party that paid for the algorithms to prioritize something for you. This can lead to conflicts and obtrusive personalization. Different digital environments serve different interests and thus capture different areas of preferences. Algorithms generalize and simplify, as they continuously filter out details that are considered irrelevant or useless. In many cases, algorithms use other people's data to fill in missing bits and pieces.

Today our algorithmic selves are beyond our control and can leave us vulnerable. A possible solution could be to have software tools and algorithms that are on our side and under our control — algorithmic guardians — capable of somehow protecting us from undesirable behavior on the part of third party algorithms. We envision algorithmic guardians as far more evolved instances of current personal assistants such Siri, Watson, etc., that, thanks to wearable and advanced human-computer interactions models enabled by pervasive computing, will be always and easily accessible for interaction. Our digital guardian will protect us from algorithmic manipulation that restricts personal freedom and will make sure that we are not stuck on repeating behavioral loops or virtual echo-chambers. It will create an adaptive information interface that is fresh and relevant. Furthermore, guardians will support us in controlling our personal data flows and deciding who can access our digital trails. For instance, with reference to the home heating example in the Data Access Control section above, our personal guardian should be able to alert us that the heating systems is still acting as if we still had the flu, should help us correct such behavior.

Our digital guardian does not need to be intelligent in the same way as humans. It needs to be smart in relation to the environments it inhabits — in relation to the other algorithms it encounters. In any case, even if algorithmic guardians (unlike third party algorithms) are user-owned and are totally under our own control, being able to understand how they work will be a priority to make them fully trustworthy. This issue, which applies to all algorithms that will govern our life — is elaborated in the next section.

Democracy Through Grey-Box Code for Algocracy

Algorithms are created by a handful of people or, in a probable near future, by other algorithms that are created by an even smaller number of people. This raises an enormous challenge for our democracies.

In the spirit of "freedom of information," publishing the algorithms towards citizens seems like an obvious thing to do. However, this doesn't make a lot of sense if the overwhelming majority of the citizens cannot even read them. In today's democracy, civil law is also produced by a handful of specialists and to use it, we rely on lawyers. Nevertheless, the civil code can be consulted by anyone who is willing to make the mental effort of digging through some heavy prose. With the current literacy level of the general public in computer science, however, there is no analogue whatsoever for public understanding of algorithms. If rules and regulations were to be expressed in code, only a tiny fraction of today's society would be able to read them.

There is compulsory need, for both citizens and governors, be able to code the rules that govern the algocracy in a format that is understandable to policy makers and to the public. We also need to make sure that we will be able to understand whether the code is actually serving the purposes it has been built for, or if it is instead bugged or hacked. In this regard, we envision three key requirements: *better programming languages, inspectability of code, and users' computing literacy.*

Concerning programming languages, we emphasize that future algorithmic governance for Smart Cities needs to be inherently distributed and mobile. Accordingly, programming for Smart Cities will require much better programming languages than currently used for distributed programming where, for instance, there is little support for the verification of the current behavior of programs. This is confirmed by the tremendous amount of middleware that exists to cover the programs' shortcomings. Yet these middleware programs do not integrate well with the host language of the code [19]. Powerful languages that allow complex distributed code to be written in a "clean" way (such as AmbientTalk [20]) have not yet made it to the mainstream. What makes a "good" language for this job? An important yardstick for measuring the quality of a language can be found in Brooks' paper on complexity in software engineering [21]. Today's programming languages put far too much emphasis on the accidental complexity of a distributed system. Languages that allow a distributed programmer to only focus on the essential complexity are still being researched at this time.

For inspectability, algorithms are often referred to as "black boxes" in that it is not apparent to the casual observer exactly how the algorithm works. Tomorrow's intelligent environments and algocracies should be based on "grey box" systems that citizens can read and tweak along various "levels" of participation. Just like there is a distinction between a constitution and normal laws, distinctions have to be made between various levels of code so that some code can be easily tweaked by direct democratic processes (à la WikiPedia), whereas other code is proverbially carved in stone. In other words, the code that runs the algocracy needs to be exposed in a "grey box" fashion, where different shades of grey will probably be needed. For instance, in the area of pervasive computing and the Internet of Things, approaches to user-level programming for configuration of smart environments, based on simple and understandable "if this then that" rules (see e.g., www.ifttt.com), and hiding more mundane programming details, go in that direction of a "grey box" approach.

In a broader perspective, societal engagement would also include building "institutions and tools that put the society in-the-loop of algorithmic systems, and allows us to program, debug, and monitor the algorithmic social contract between humans and governance algorithms" [22]. The need for transparency, accountability, and explainability for the increasingly prevalent AI "blackboxes" has been noted in [23], where a layered model involving technical, ethical, legal, and social aspects needs to be taken into account.

In parallel with the development of an understandable "gray box" approach to programming, we need to solve one of the main factors hampering efforts for a healthy algocracy. This is the need for citizens to have at least a basic literacy in computing, and — if they are not able to program — they should be capable of judging the actions and the quality of the programs that govern them (at least when exposed in their "gray box" form). Unfortunately, computer science as a basic scientific field is absent in the high school systems of most countries. A notable exception is the U.K. where "Computing" is part of the high school curriculum since 2015. Even for people who will never program in their entire lives, a good basic understanding of what is programming, and what is an algorithm, is necessary for being a citizen in the algocracy!

Humans in the Loop and the Wisdom of Many

An algorithm is weightless and only worth the weight people put in it, so that some degree of safety from the potential dangers of algorithmic governance can come from the "wisdom of the crowd." There is a need for users to be able to provide feedback to the system in a forum, or through a mechanism for collectively commenting on the algorithm's performance, so that problems can be identified and signaled. The same pervasive sensing technologies that feed the algorithms with data can be exploited by users to monitor the environment and signal (and share information about) problems caused by existing Smart City algorithms. For instance, the fact that home heating systems are biased so that they do not meet user needs and only act towards some municipality goals, can be discovered by a multitude of users (or by their algorithmic guardians) by accessing sensor and actuator data. Eventually the users can then make the facts emerge to global awareness. Indeed, it should be a general goal for governments (at all levels, from national to municipal), whenever they start relying on algorithms to control cities and make decisions on our behalf, to involve human citizens in the loop at the highest levels of the participation ladder [3], i.e., as partners, in delegating authority, and as co-managers of the system or algorithm.

A different possible form of the wisdom of the crowd can be the wisdom of the crowd of algorithms, enabled by the existence of a plurality of algorithms and systems devoted to govern the same concern. In essence, human decisions might be based on some aggregation of the inputs of a number of preferably independent algorithms. That is, we can take the principle of the "wisdom of the many" to systems. However, when algorithms deal with ethical choices, the involvement of humans may become necessary. Should there be no time to involve a human in the decision-making, one can consider that some algorithmic decisions are premade by a human in advance so that human accountability is retained, and automatically adapted to the specific context once these decisions have to become actions.

Lastly, when there are systematic failures, there needs to be a way to "pull the plug." When an algorithm it is found to deliver unfavorable outcomes or cause problems, humans must be able to stop using the system. As a simple example, a user (as it is already indeed the case in our homes) should always be able to turn off an automatic home heating system and operate the system manually. The mechanism for control over these systems, and the way to turn them off, needs to be obvious [24]. However, for critical systems, the look and feel of the control switches needs to be different, so that unintentional, potentially fatal, mistakes are not made. As related in [24], control-room operators in a nuclear power plant found that similar-looking knobs could lead to a disastrous outcome, hence beer-keg handles were placed over them. Putting humans in ultimate control of algorithms would seem sensible, but is not without its own issues. For example, a human cannot simply switch off an autonomous vehicle when s/he thinks it is not performing up to its requirements — there needs to be a way to deliver control back to humans safely, and once in control, for humans to safely control the algorithm.

From Context-Awareness to Context-Control-Awareness

Pervasive computing technologies enable algorithms to make decisions that are context-aware, i.e., adapted to the context in which they operate (the already mentioned personalization being a specific form of context-awareness). Context-awareness has been studied since the 1990s in the area of pervasive computing [25], [26]. The last decade has seen dramatic progress in automatic recognition of context (including place — outdoors and indoors, human activity, habit, preference, and available energy and resources, etc.) and the self-adaptation of pervasive computing devices to the learned context.

Given the possible perils of algorithmic governance, one possible research direction is *"context-control-aware"* systems. That is, algorithms that can override control are not given full access for adapting the devices under their influence but are, instead, given a shared access control with a network of socially connected devices, with humans as co-decision makers, for shared governance [27]. This allows pervasive computing systems to be more "considerate," as they are not only aware of their contexts (and input to the systems), but also of how their influence and control (and the output of the systems) can bring unintended consequences. Such shared control should also account for safety, security, and privacy that — although extensively researched so far mostly in a separate silo — have yet to be integrated in context-control-aware systems for safer smart environments.

The issue of control (also related to the previously mentioned issue of "pulling the plug") involves that of making Smart Cities and environments really usable. This alludes to Steve Krug's attributes of usability of an interface [28]: useful, learnable, memorable, desirable, and delightful. In particular, the studied contexts in context-control-aware computing for Smart Cities and environments should also be provided with interfaces allowing citizens and end users to voice any discomfort and displeasure with the systems. This would enable higher levels of interactivity with the governing algorithms, and enable these systems to, e.g., "reverse action," "pull back," and activate "dumb mode" when required. We need Smart Cities to not just be "efficient." We also need to make it possible for citizens to be more interactive with the governing of cities, and vice versa [29].

There is a need for balance between a system that is too obtrusive to be useful, in which the user is too often involved, and a system that is too autonomous so that algorithmic regulation becomes real. A question is, can an algorithm be designed to compute this balance? Can an algorithm "solve" the problem of algorithmic dominance or governance, or of being context-aware, including an algorithm being aware of itself being too controlling? This is where potentially the pervasive computing community, in a close collaboration with other fields in this truly complex multidisciplinary issue, can contribute a solution.

Protecting the Individual Citizen from the Algorithm

The obvious advantage of the increasing availability of pervasive computing infrastructures is that they will make our lives much easier. However, this evolution also carries potential risks for individuals and for society as a whole. By blindly accepting the deployment of algorithms that run our society, we could become a "dumber" society or even lose control. In this context, "algocracy" may be nontrivial to reconcile with democracy. Dealing with these issues will require deploying a system of societal apparatuses to protect the individual citizen against the running code and/or against potentially malicious use by individuals of the data that is collected and produced by that code.

In this article, we have explored five avenues. 1) First, there is the obvious attention to data access control. Beside traditional privacy and security concerns, the notion of behavioral privacy will be equally important, as it will be the possibility for understanding how the exploitation of our personal data affects algorithms' behavior. 2) Part of the solution may be to proactively chaperone the ongoing activities of the algocracy with "algorithmic guardians" that can represent and defend us in the algorithmic world. 3) A less trivial challenge lies in making citizens aware of the code that runs their algocracy, and empowering them in novel democratic procedures that will be used to manage that code. 4) We should never give up the possibility for humans to "pull the plug" or to insist on the wisdom of a crowd of (preferably independent) algorithms. 5) Finally, context awareness could be used to "sandbox" the power of certain algorithms in certain contexts and to provide the meta technology to activate and deactivate the sandboxing based on a citizen's expression of discomfort.

We agree with [30] that ethical considerations must be central to new algorithms we will create in the future.

Author Information

Franco Zambonelli is with the Università di Modena e Reggio Emilia, Italy. Email: franco.zambonelli@unimore.it.

Flora Salim is with RMIT University, Australia. Email: flora.salim@rmit.edu.au.

Seng W. Loke is with Deakin University, Australia. Email: seng.loke@deakin.edu.au.

Wolfgang De Meuter is with Vrije Universiteit, Brussels, Belgium. Email: wdmeuter@vub.ac.be.

Salil Kanhere is with the University of New South Wales, Australia. Email: salil.kanhere@unsw.edu.au.

References

[1] F. Zambonelli, "Toward sociotechnical urban superorganisms," *Computer*, vol. 45, no. 8, pp. 76-78, Aug. 2012.

[2] A. Schmidt, M. Langheinrich, and K. Kersting, "Perception beyond the Here and Now," *Computer*, vol. 44, no. 2, pp. 86-88, Feb. 2011.

[3] S.R. Arnstein, "A ladder of citizen participation," *J. American Planning Assoc.*, vol. 35, no. 4, pp. 216-224, July 1969.

[4] N. Rodrigues, "Algorithmic governmentality, Smart Cities and spatial justice," *justice spatiale - spatial justice,* Université Paris Ouest Nanterre La Défense, UMR LAVUE 7218, Laboratoire Mosaïques, Liberty, Equality, IT, 10, 2016; http://www.jssj.org/article/gouvernementalite-algorithmique-smartcities-et-justice-spatiale/.

[5] N. Just and M. Latzer, "Governance by algorithms: Reality construction by algorithmic selection on the Internet," *Media, Culture & Society*, vol, 39, no. 2, pp. 238 - 258, Apr. 2016.

[6] D. Doneda and V.A.F. Almeida, "What is algorithm governance?," *IEEE Internet Computing*, vol. 20 no. 4, pp. 60-63, 2016.

[7] L. DeNardis and A.M. Hackl, "Internet governance by social media platforms," *Telecommunication Policies*, vol. 38, no. 9, pp. 761-770, Oct. 2015.

[8] C. Borean, R. Giannantonio, M. Mamei, D. Mana, A. Sassi, and F. Zambonelli, "Urban crowd steering: An overview," in *Proc. Int. Conf. Internet and Distributed Computing Systems, Lecture Notes in Computer Science*, no. 9258, 2015, pp. 143-154.

[9] Maslow, "Higher and lower needs," *J. Psychology: Interdisciplinary and Applied*, vol. 25, no. 2, 1948.

[10] S. Hajian, F. Bonchi, and C. Castillo, "Algorithmic bias: From discrimination discovery to fairness-aware data mining," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD '16).* New York, NY: ACM, 2016, pp. 2125-2126.

[11] S. Wachter-Boettcher, *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech.*Norton, 2017, p. 10.

[12] Rouvroy and B. Stiegler. "The digital regime of truth: From the algorithmic governmentality to a new rule of law," *La Deleuziana – On Line J. Philosophy*, no. 3, pp. 6-29, 2016.

[13] A. Lee, "Geoffrey Hinton, the 'Godfather' of Deep Learning, on AlphaGo," *MacLeans*, Mar. 18, 2016; http://www.macleans.ca/society/science/the-meaning-of-alphago-the-ai-program-that-beat-a-go-champ/.

[14] K. Crawford, "Can an algorithm be agonistic? Ten scenes from life in calculated publics," *Science Technology Human Values*, p. 162243915589635, Jun. 2015.

[15] "Open letter on autonomous weapons," *FLI - Future of Life Institute*, Jul. 28, 2015; http://futureoflife.org/open-letter-autonomous-weapons/, accessed Jul. 18, 2016.

[16] J.F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science*, vol. 352, no. 6293, pp. 1573-1576, Jun. 2016.

[17] H. Choi, S. Chakraborty, and M. Srivastava, "Design and evaluation of SensorSafe: A framework for achieving behavioral privacy in sharing personal sensory information," in *Proc. TrustCom* (Liverpool, U.K.), Jun. 2012.

[18] P. McDaniel and S. McLaughlin, "Security and privacy challenges in Smart Grid," *IEEE Security and Privacy*, vol. 7, no. 3, pp. 75-77, May-Jun. 2009.

[19] J-P. Briot, R.Guerraoui, and K-P. Lohr, "Concurrency and distribution in object-oriented programming," *ACM Computing Surveys*, vol. 30, no. 3, pp. 291-329, Sept. 1998.

[20] T.V. Cutsem, S. Mostinckx, and W.D. Meuter, "Linguistic symbiosis between event loop actors and threads," *Computer Languages, Systems and Structures*, vol. 35, no. 1, pp. 80-98, 2009.

[21] F.P. Brooks, "No silver bullet: Essence and accidents of software engineering," *Computer*, vol. 20, no. 4, pp. 10-19, Apr. 1987.

[22] I. Rahwan, "Society-in-the-loop: Programming the algorithmic social contract," *Ethics and Information Technology*, vol. 2, no. 2, pp. 1572-8439, 2017.

[23] U. Gasser and V.A.F. Almeida, "A layered model for AI governance," *IEEE Internet Computing*, vol. 21, no. 6, pp. 58-62, Nov./Dec. 2017.

[24] J.L. Seminara, W.R. Gonzalez, and S.O. Parsons, "Human factors review of nuclear power plant control room design," Report no. EPRI NP-309, Electric Power Research Inst., Palo Alto, CA, 1977.

[25] B. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Proc. Workshop on Mobile Computing Systems and Applications*, 1994, pp. 85–90.

[26] G.D. Abowd, A.K. Dey, P.J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *Handheld and Ubiquitous Computing*, H.-W. Gellersen, Ed. Berlin-Heidelberg, Germany: Springer, 1999, pp. 304–307.

[27] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir, "The future of human-in-the-loop cyberphysical systems," *Computer*, vol. 46, no. 1, pp. 36–45, Jan. 2013.

[28] S. Krug, *Don't Make Me Think: A Common Sense Approach to Web Usability*. India: Pearson, 2005.

[29] F. Salim and U. Haque, "Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things," *Int. J. Human-Computer Studies*, vol. 81, pp. 31–48, Sept. 2015.

[30] D. Bianchini and I. Avila, "Smart Cities and their smart decisions: Ethical considerations," *IEEE Technology and Society Mag.*, vol. 33, no. 1, pp. 34-40, Spr. 2014.

[31] R.N. Charette "Automated to death," *IEEE Spectrum*, Dec. 15, 2009; https://spectrum.ieee.org/computing/software/automated-to-death.

Last Word

Christine Perakslis

Digital Empowerment and Socio-Political Stability

Social media (SM) usage is increasing across the globe. Of the 7.6 billion people populating earth, 4 billion are believed to be Internet users. Over 3 billion are SM users, representing over 40% global penetration [1], [2].

In this issue of *IEEE Technology and Society Magazine*, we contemplate SM; we can postulate effects on political, economic, socio-cultural, technological, legal, and environmental (PESTLE) factors. We reviewed how social networking sites (SNS) are channels used to socio-politically raise awareness and mobilize people during elections. We reviewed opinions shared by females who face strict gender segregation rules.

We brought special focus to the Middle East, the transcontinental region with an estimated 130 million active SM users (up almost 40% from 2017) and 164 million Internet users (or 65% penetration) [1], [2]. Our authors confirmed findings from previous studies [3]: SM users in the Middle East describe a lack of freedom of expression, and worry about legal consequences when creating and sharing content. Some users can face fines and prison sentences if posts are interpreted as critical or insulting, or if photographs or videos of others are posted without consent [4].

Some clerics, or others in authority in this region, have blamed SM for uprisings. Therefore, some governments block, restrict, or prohibit popular sites and services. Yet, SM are not the causes of discontent or disaffection. SM are channels of communication. Discontent and disaffection are upstream, flowing downstream into many various tributaries for communication; tributaries can be online and offline channels.

Yet, undeniably, SM continue to prove powerful. Communication is a formidable tool for empowerment. Citizens can use SM to communicate to influence large audiences, to build group identity and unity, and to expose abuses. SM can give voice to the voiceless. People can become empowered.

In an interconnected world, the autocratic will continue to face challenges with the digital. Savvy users persist to find methods to bypass communication suppression, much like the 46% of survey respondents in Arab regions reporting multiple accounts on a single SNS platform [5].

With all the suppression efforts aimed at avoiding political instability, perhaps the autocratic misjudge the empowerment borne through SM. Particularly, because empowerment/civic participation is believed to be one of the four necessary conditions required to achieve sustainable political stability [6].

Digital Object Identifier 10.1109/MTS.2018.2837539
Date of publication: 31 May 2018

Author Information

Christine Perakslis is Associate Professor in the MBA Program, College of Management, Johnson & Wales University, Providence, RI. Email: christine.perakslis@jwu.edu.

References

[1] S. Kemp, "Digital in 2018," wearesocial.com, Jan. 2018; https://wearesocial.com/blog/2018/01/global-digital-report-2018, retrieved Feb. 1, 2018.

[2] T. Elmasry et al., *Digital Middle East: Transforming the region into a leading digital economy.* Digital McKinsey, Oct. 2016; https://www.mckinsey.com/global-themes/middle-east-and-africa, retrieved Feb. 1, 2018.

[3] Y. Al-Saggaf, "Saudi females on Facebook: An ethnographic study," *Int. J. Emerging Technologies and Society*, vol. 9, pp. 1-19, 2011.

[4] Staff writer, "GCC national jailed for 10 years, fined for Twitter violations," *arabianbusiness.com*, Dec. 2017; http://www.arabianbusiness.com/media/386492-gcc-national-jailed-for-10-years-fined-for-twitter-violations.

[5] F. Salim, "Social media and the Internet of Things: Towards data-driven policymaking in the Arab World: Potential, limits and concern," Mohammed Bin Rashid School of Government, Dubai, 2017; http://www.mbrsg.ae/getattachment/1383b88a-6eb9-476a-bae4-61903688099b/Arab-Social-Media-Report-2017, retrieved Jan. 20, 2018.

[6] U.S. Institute of Peace, *Guiding Principles for Stabilization and Reconstruction*, U.S. Army Peacekeeping and Stability Operations Institute, 2009; https://www.usip.org/sites/default/files/guiding_principles_full.pdf, retrieved Feb. 10, 2018.

SSIT Fundraising Campaign

Influence the Direction of Technology!

IEEE

IEEE SSIT SOCIETY ON SOCIAL IMPLICATIONS OF TECHNOLOGY

- SSIT brings together interdisciplinary communities to explore the evolution of, and inform our understanding of
  - Sustainable Development & Humanitarian Technology
  - Ethics and Human Values
  - Technology Access
  - Societal Impact of Technological Innovation
  - Protecting the Planet
- We work *within* the technology community, but draw our ideas from practitioners, researchers, inventors, policy makers, professionals and students across many fields
- We influence the direction of technology adoption and development, from awareness and concepts, to standards and professional practice

To seize opportunities* in 2018 we need $50,000.

With *your* help, these programs *will* make a difference!

*Funds will be invested in further strengthening and expanding:

- Chapter, Young Professional and Student Activities
- Conference, Distinguished Lecturer (DL) and Standards Programs
- Online Publishing and Education Programs

Contact us for details.

Ways to Contribute

- Donate to SSIT online at https://ieeefoundation.org/ieee_ssit
- You can make a gift to SSIT in honor or memory of someone who has touched your life

Donations to SSIT are managed by the IEEE Foundation, the philanthropic arm of IEEE. IEEE and the IEEE Foundation are U.S. 501(c)3 non-profit organizations. For more information contact: donate@ieee.org or +1 732 465 5871.

www.TechnologyandSociety.org