Ramona Pringle   Katina Michael   MG Michael

I n November 2015, as editors, we started sorting through submissions for a special collection of articles examining the paradox of technological potential. Separately, we published a call to the public at large, asking *"What comes to mind, when you think about technology, and unintended consequences?"*

Then using a smart phone, we documented handwritten responses from people whose paths we crossed, collecting submissions from individuals in both hemispheres. While geographically the contributors could not have been further apart, psychologically, their responses echoed one another in a resonant, prescient way. In this way we were able to physically capture the thoughts of a variety of participants, and share them with our co-editors across the globe in near real time thanks to the marvels of technology.

Simultaneously, the responses we gathered noted some recurring and challenging concerns. We heard concerns about time, presence, and disconnection. Many of these responses have been archived in a photographic collection [1].

Unanticipated consequences is a sociological concept emphasized by Robert K. Merton [2]. Although later writers have interchanged the

Unintended Consequences of Living with AI

The Paradox of Technological Potential—Part II

"**Tons of information but not enough understanding**"

"**Words cannot be swallowed once spoken.**"

"**Too many distractions.**"

"**…That unfortunate moment when you hit reply all.**"

notion of *unintended consequences* with *unanticipated consequences*, the two phrases are subtly but significantly different, despite being deeply connected [3]. In general, unintended consequences refer to those not intended by a purposeful action. Unanticipated consequences are those with outcomes that were not those that were foreseen. It follows then that an unintended consequence might/might not have been anticipated (Table 1). It is also important to state that unintended consequences can have positive, negative, or even perverse [4] impacts on individuals, groups of people, or society at large (Figure 1).

Some sociology of science scholars refer to this phenomenon more generally as the *law of unintended consequences*, defined as the actions of people, groups, organizations, or governments that may be either anticipated or unanticipated in effect. It is a set of results that was not intended as an outcome, but happened regardless. The vast majority of people consider unintended consequences to be disadvantageous, counterproductive,

Consequences	Anticipated	Unanticipated
TABLE 1. (Un)Anticipated (Un)Intended Consequences.		
Intended	"Known knowns." Positive, negative or perverse consequences on society.	Invalid category. A consequence cannot be intended but unanticipated. For this reason this is often referred to as the "Unknown knowns." That which we intentionally refuse to acknowledge that we know.
Unintended	"Known unknowns." Opportunity to curb negative or perverse consequence(s) before it/they happen. Equally, unintended consequences that are anticipated and beneficial and can have positive impacts on society.	"Unknown unknowns." Usually these unanticipated and unintended consequences have significant negative consequences on society. Very few of these have positive impacts because of their unpredictability.

fraudulent, or at times detrimental and even dangerous.

It is a paradox that today we think of disruption as an intentional action intended to trigger market forces and spur adoption of new technologies. Contrarily, unintended consequences do not have a purposeful intentionality about them. Quite often, the creator of an innovation does not attempt to steer the adoption and use of their product in the direction that eventually results. This leaves us only to ponder "after the fact" that the consequences were unintended by the creator.

In fact, there would be nothing stopping the rest of society from speculating as to what some of these unintended consequences might be prior to commercialization.

The former United States Secretary of Defense, Donald Rumsfeld, brought the idea of "known knowns" to prominence when he answered a question at a U.S. Department of Defense (DoD) news briefing on February 12, 2002, linking the distribution of weapons of mass destruction (WMD) with the Iraqi government (5). He constructed three contexts, saying:

"Reports that say that something hasn't happened are always interesting to me, because as we know:

1) There are *known knowns*; there are things we know we know.
2) We also know there are *known unknowns*; that is to say we know there are some things we do not know.
3) But there are also *unknown unknowns* — the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones" [6] *(numbering added).*

The idea of *unknown unknowns* was created by Joseph Luft and Harrington Ingham with the development of the Johari window in 1955 (7), which was centered on understanding the self and our relations with others. It was a concept that was widely used in NASA with respect to decision-making and risk.

While Rumsfeld referred to three categories, some philosophers such as Slavoj Žižek have proposed a fourth category of the "known unknowns." Žižek writes that these are: "the disavowed beliefs, suppositions and obscene practices we pretend not to know about, even though they form the background of our public values" (8). German sociologists Daase and Kessler (9) point out that while Rumsfeld emphasized the cognitive frame for political

practice in *what we know, what we do not know,* and *what we cannot know,* he failed to address *what we do not like to know* (9).

If we take these underlying concepts and integrate them to the technological realm as in Table 1 we get:
1) The *known knowns* are those things that are anticipated consequences that history has taught us might happen in a given context. For example, some members of society may not be able to afford new technologies and so a digital divide between the haves and have-nots sets in with the increasing number of consumer technologies introduced into the market (i.e., difference between mobile phone, smart phone, smart watch, microchip implants) (10).
2) The *known unknowns* are again things that can be predicted, but often by applying good judgement and common sense principles about the possibilities. For example, the introduction of consumer products like the *Hello Barbie, Alexa, DropCam,* and *Nest* devices that have been deployed with significant social, privacy, and security problems that will inevitably create some benefit, but even greater drawbacks in unknown effects. Some would go so far as to say that these types of consumer products may have perverse impacts on

particular groups like children or the mentally ill.

3) The *unknown unknowns* are completely unexpected, and are for the most part unpredictable because no evidence exists to identify particular risk factors or even attributes linking them to a given phenomenon. As technologies increase in complexity, we should expect a greater number of unknown unknowns, with a greater severity of consequence. So-called "humane robots" that are used in assisted living contexts to aid the elderly (e.g., to get dressed, remember to take their tablets, provide fall-down alert support and more) presently have a lot of unknown unknowns attached to them (11), though equally they may have benefits that cannot be disputed.

4) The fourth and final category is the *unknown knowns*. This category is somewhat of an oxymoron. It cannot be and yet it is. While creators of new technologies are likely in many ways to be in the best position to critique their own creations, most innovators openly say that they have no idea how their creations will be used by society, and that we cannot prejudge ethics. We think of technologies such as gaming apps for the smartphone that are built to encourage return visits with stickiness features. Yet some software developers will never admit to triggering addictions in members of the populace prone to addictive behaviors (12).

Such as in the fourth category above, some technologists in the areas of artificial intelligence and robotics tend to shrug off plausible anticipatory outcomes, knowing full well what the outcomes could mean for society at large in the contexts of social and behavioral problems, or even privacy and security.
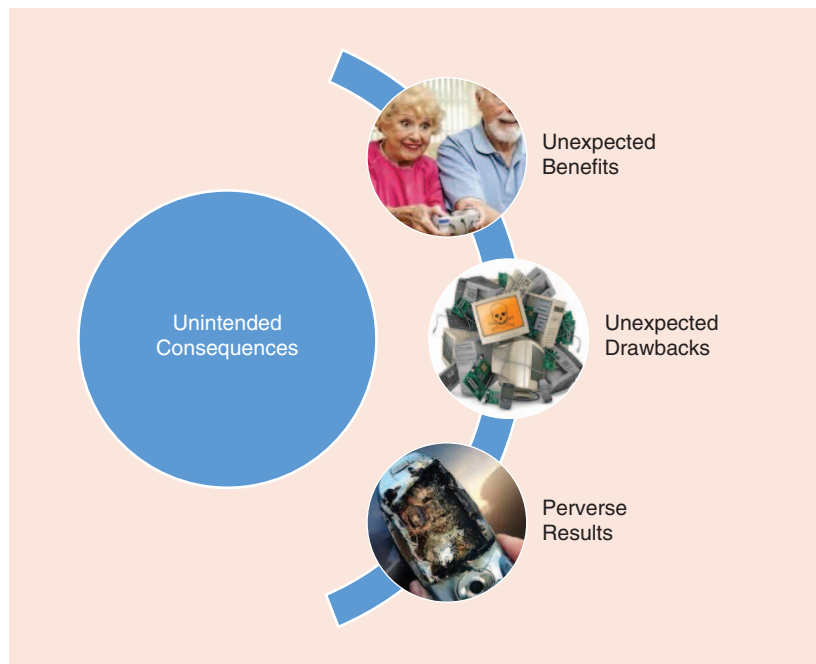
FIGURE 1. The social implications of unintended consequences.

Propelled by a yearning to create, develop, and deploy, and to be the first movers, they may a) play down their creations and the impact they will have; b) refuse to acknowledge that anything might go wrong; or c) when they do admit a potential risk, they still ask the rest of us to go down that bleak road with them. At the point of development and potential launch of a new product, the momentum is likely to be such that we consumers do follow after them. We follow by purchasing the new systems or products, or by asking few questions. We might fail then to ask questions, say about how a new technology might further distance us from reality, or about how it might distance us further from our human relationships (13).

At the moment of writing this editorial (October 21, 2016), nearly half of the Internet in the United States is down. A massive distributed denial of service (DDoS) attack targeted a company that functions essentially as a switchboard for the U.S. Inter-

net, translating human-facing web addresses to the numerical mode of communication used by computers. This attack is different from DDoS attacks we have seen in the past. It is larger and more powerful. By targeting a company that manages the infrastructure of the Internet, the attack has impacted several mainstream websites, as opposed to a single corporation or organization.

These types of attacks are actually *known knowns.* That is, we can have a constant expectation of such attacks on the Internet. We can predict that these types of attacks will only increase in number exponentially, and that they will have ever greater global economic impact.

But then there are also *unknown knowns* related to this latest attack. Unknown knowns in this case would be vulnerabilities embedded into a world built on the Internet of Things (IoT). This October 16 attack relied on infected smart devices — networked items such as security cameras, home routers, and baby

monitors — to direct traffic towards the target in an effort to overwhelm and compromise its servers. These distributed devices form a botnet, a network of Internet-connected private computers, infected with malicious software and controlled as a group. In this case, these IoT smart devices would have been in people's homes and businesses, and infected without their owner's knowledge.

As the blog *KrebsonSecurity* notes, these "inexpensive, mass-produced IoT devices are essentially unfixable, and will remain a danger to others unless and until they are completely unplugged from the Internet." Of course, unplugging these devices is an unlikely scenario, given that the owners of these devices presumably do not even know that their cameras or baby monitors are infected [14]. The phenomenon of smart devices, which are actually extremely dumb when it comes to security, is a timely example of *unknown knowns*. As developers race to get these impressive tools to market, they overlook key security and privacy considerations—and ask that the public turn a blind eye as well.

While we consider that which is known and that which is unknown, it is important to define technological potential within a context of sustainability. Competence, skills, and effectiveness in R&D activities, as well as scientific-industrial relations within the economy must be considered [15]. We want to glean the future and consider what might aid users in the longer term [16]. We should seek new ways to meet challenges in terms of materials, processing capabilities, product functionality and use-value. In this manner the idea of "potential" is steeped in technological forecasting. But realizing technological potential cannot be accomplished in a vacuum. The forecaster must weigh things like public opinion and pricing pressures, "the possibility of changes in institutional resistances, and the probable future marginal preferences of the society" [17]. Understanding target markets with respect to new innovations is critical, and designing technologies with these markets in mind, and how they might exploit a future product is critical. Entrepreneurs who embrace technological potential at the core of their efforts are usually more optimistic than pessimistic about the social impact of their investments.

Advancement of new technologies will be essential to solving many of the most complex challenges of the 21st century—from diseases to climate change. We need only look at research in the field of alternative renewable energy sources, e.g., in various types of fuel cells, to be inspired and to consider the potential possibilities.

But what happens when a given technology is adopted for perverse ends? Or when technology is knowingly deployed to propel unhealthy practices beyond theoretical limits of its use? What if the consequences of a given technology are contrarily subsuming healthy human emotions in adults and children alike? Who then is responsible for that innovation and for the asymmetric impact it will inevitably have on many people's lives?

These issues are front and center, or should be, in the development of artificially intelligent agents, as we are essentially creating a new technological "species," an undertaking ripe with controversy and complexity.

Our original Call for Papers for this topic, for *IEEE Potentials* magazine, examining the Paradox of Technological Potential (PTP), was answered with so many strong articles on the multifaceted relationship between human and machine, and the complexity of living with AI, that we have additionally chosen to also devote this special section of *IEEE Technology and Society Magazine* to this important subject.

Assisted living machines, for example, that were built to aid humans, could be rendered killing machines if misused or misapplied to state enemies. What was once the province of science fiction visions is now within the realm of reality. The paradox is in the contradiction of the potential, which can be used for both good and bad, but not for both at the very same time in a given implementation [18].

Part I of this project was published in *IEEE Potentials* in September/October 2016 as a Special Issue. Much of that special issue had to do with the relationship between the veillances (*sur-, sous-, uber-*), evidence-based /intelligence-led policing and counter-strategies to mass surveillance, predictive profiling, and finally privacy and security by design.

In this special section of *IEEE Technology and Society Magazine* [20], our focus is more on a futurist vision of the technological potential of AI, with a check on unintended consequences, and a specific focus on the complexities of life during the coming of age of artificial intelligences. Our question is, in the development of artificial intelligence, what is driving us? Are visions of science fiction propeling us to a Silicon Valley dreamworld filled with technologies that we well know will be dystopic?

Edward Tenner, who has deeply studied the cultural aspects of technological change, has noted that although our capabilities and technology have been expanding geometrically, our "ability to model their long-term behavior" has not kept pace with the change [21]. Tenner believes that one of the problems of our time is how we will close the gap between capabilities and foresight.

In closing his March 2011 TEDtalk, he emphasized that we are living in a time of unexpected possibilities and that the secret to our future may well be to take a "really positive view" of unintended consequences in going forward.

We as guest editors of this *Technology and Society Magazine* special section prefer to take the approach of "cautious optimism," whereby we can steer technologies toward sustainable causes, and then expect at least some positive return. Unknown unknowns may not be the biggest problem we are facing, but deliberately covering up the so-named unknown knowns.

We would like to thank the authors and reviewers for their contributions. In this section we have incorporated perspectives that explore methodologies that could be used to build a better future, and that investigate the potential of new technologies like wearables. We consider the impacts of artificial intelligence through science fiction, looking at the way that the future field of robotics might apply to everyday people, including the implications for everyday citizenry. We've also included a fiction piece on the possibilities of crossing the human evolutionary gap, and even an original interview with a humanoid robot, a sure sign of the times we live in. Augmenting this material, the editor has included relevant articles on the future relations between humans and artificial intelligence, on help for those subject to Internet externalities like cybersex addiction and online Internet addiction, on how to reclaim conversation, and more.

Author Information

Ramona Pringle is an assistant professor in the RTA School of Media at Ryerson University, Toronto, and Creative Director of the Transmedia Zone, an incubator for the future of media. Email: ramona.pringle@ryerson.ca.

Katina Michael is a professor in the Faculty of Engineering and Information Sciences at the University of Wollongong, NSW, Australia. Email: katina@uow.edu.au.

MG Michael, is an honorary associate professor at the School of Computing and Information Technology at the University of Wollongong, NSW, Australia. Email: mgm@uow.edu.au.

Acknowledgment

References

[1] K. Michael, *Unintended Consequences 1–100*, Sapphire Coast, Dec. 2015; http://www.katinamichael.com/call-for-papers/2016/1/14/unintended-consequences-1-100-artwork, accessed Oct. 11, 2016.

[2] R.K. Merton, "The unanticipated consequences of purposive social action," *American Sociological Rev.*, vol. 1, no. 6, pp. 894–904, Dec. 1936.

[3] F. de Zwart, "Unintended but not unanticipated consequences," *Theory and Society*, Vol. 44, Iss. 3, 2015, pp. 283–297.

[4] R. Norton, "Unintended consequences" in *The Concise Encyclopedia of Economics*, 2nd ed.; http://www.econlib.org/library/Enc/UnintendedConsequences.html, accessed 14 Oct. 2016.

[5] D. Rumsfeld, *Known and Unknown: A Memoir.* New York, NY: Sentinel, 2011; https://books.google.com/books?id=_wlcpxMOjD4C&pg=PT15, accessed 27 Oct. 2016.

[6] Defense.gov. DoD News Briefing — Secretary Rumsfeld and Gen. Myers, *United States Department of Defense*, http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636, 12 February 2012.

[7] J. Luft and H. Ingham, "The Johari window, a graphic model of interpersonal awareness," *Proceedings of the western training laboratory in group development.* Los Angeles: University of California, Los Angeles, 1955.

[8] S. Žižek, "What Rumsfeld doesn't know that he knows about Abu Ghraib," *ŽIŽEK.UK*; http://zizek.uk/what-rumsfelddoesnt-know-that-he-knows-about-abu-ghraib/, accessed 27 Oct. 2016.

[9] C. Daase and O. Kessler, "Knowns and unknowns in the `War on Terror': Uncertainty and the political construction of danger," *Security Dialogue*, vol. 38, no. 4, pp. 411–434, Dec. 2007.

[10] K. Michael and MG Michael, *Innovative Automatic Identification and Location-Based Services: From Bar Codes to Chip Implants.* Hershey, PA: Information Science Reference, 2009, p. 416.

[11] R. Pringle, "Are robots coming for our jobs?," *CBC News*, Sept. 19, 2016; http://www.cbc.ca/news/technology/are-robots-coming-for-our-jobs-1.3765986, accessed 27 Oct. 2016.

[12] I. Leslie, "The scientists who make apps addictive," *The Economist*, Oct./Nov. 2016; https://www.1843magazine.com/features/the-scientists-who-makeapps-addictive, accessed 27 Oct. 2016.

[13] MG Michael and K. Michael, "Resistance is not futile, *nil desperandum*," *IEEE Technology and Society Mag.*, vol. 34, no. 3, pp. 10–13, Sept. 2015.

[14] Krebs, "Hacked cameras, DVRs powered today's massive Internet outage," *KrebsonSecurity*, Oct. 21, 2016; https://krebsonsecurity.com/2016/10/hackedcameras-dvrs-powered-todays-massiveinternet-outage/, accessed 27 Oct. 2016.

[15] R. Wisła, "Gender and industrial creativity in Poland," in *Contemporary Global Perspectives on Gender Economics.* Hershey, PA: IGI, 2015, pp. 230–243.

[16] H.-J. Schellnhuber, "Technologies and their sustainable potential," *World in Transition: Towards Sustainable Energy Systems*, ch.3, 2004, p. 43.

[17] J. B. Quinn, "Technological forecasting," *Harvard Business Rev.*, Mar. 1967; https://hbr.org/1967/03/technologicalforecasting, accessed 21 Oct. 2016.

[18] GACGC. "World in transition: Strategies for managing global environmental risks, *German Advisory Council on Global Change*, Annual Report, 1998; http://www.wbgu.de/fileadmin/templates/dateien/veroeffentlichungen/hauptgutachten/jg1998/wbgu_jg1998_engl.pdf, accessed 19 Oct. 2016.

[19] Interviews with select authors from the special issue of *IEEE Potentials* Vol. 35, No. 5, 2016 edited by Ramona Pringle, Katina Michael and M.G. Michael can be found here: https://www.youtube.com/playlist?list=PLg4RQrKJML2tNe384z3fhkjCWmOe-4e-M

[20] R. Pringle, K. Michael, and M.G. Michael, Eds., "Interviews with select authors from special issue of *IEEE Technology and Society Magazine*, vol. 35, no. 4, 2016; https://www.youtube.com/playlist?list=PLg4RQrKJML2umWLuGikxVI19A857Dcws9.

[21] E. Tenner, "Unintended consequences," *TED.com*, Mar. 2011; https://www.ted.com/talks/edward_tenner_unintended_consequences/transcript?language=en#t-658000, accessed 13 Oct. 2016.

TS