Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Surviving a technological future: Technological proliferation and modes of discovery

Chris William Callaghan

Knowledge and Information Economics/Human Resources Research Agency, School of Economic and Business Sciences, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein, 2000, South Africa

ARTICLE INFO

Keywords:
Futures studies
Theory development
Openness
Power
Technological threats
Technological proliferation

ABSTRACT

Certain future scenarios of technological change are dystopian in their predictions. Fewer are optimistic. Taking a pragmatic stance, this paper seeks to identify certain key threats associated with the proliferation of dangerous technologies, giving voice to those in the literature on different sides of the debate. Novel literature is considered that suggests that innovations in the discovery, or research process itself, may hold the key to developing certain collaborative capabilities that can amplify collective intelligence. These capabilities are discussed together with their potential to meet the challenges associated with the proliferation of dangerous technologies. Testable propositions are derived from literature, and four technological scenarios are developed for analysis. Certain key challenges are identified and discussed in relation to each of the technological scenarios. In doing so, what are hopefully useful insights are derived for how changes can be made in the present to help avoid meeting the fates described by certain of these scenarios.

1. Introduction

“Almost all the problems we face nowadays are complex, interconnected, contradictory, located in an uncertain environment and embedded in landscapes that are rapidly changing” according to Sardar (2010:183). Over and above the use of nuclear or biological weapons, potential risks are associated with emergent technologies such as artificial intelligence (AI), biotechnology, geoengineering, and nanotechnology (Baum, 2015). Unlike the threats of nuclear, biological or chemical weapons of mass destruction, however, novel technologies such as genetics, nanotechnology and robotics (GNR) do not require large-scale activities to pose threats to humankind, but only require knowledge, posing the threat of knowledge-enabled mass destruction, “amplified by the power of self-replication” (Joy, 2000:1).

How does one address such threats? To do so, a choice must be made in terms of what scientific methodology to use, and a definition is required, as to what it means to ‘address’ such threats. Given the almost unimaginable uncertainties associated with technological advancement and its consequences (Vinge, 1993; Szerszynski, Kearnes, Macnaghten, Owen, & Stilgoe, 2013; Bostrom, 2017; Tegmark, 2017), and therefore the need for responsible innovation (Grunwald, 2011; Stilgoe, Owen, & Macnaghten, 2013), this paper takes recourse to the approach of future studies to develop a theoretical framework relating to how to prepare for such scenarios. In doing so, this paper also seeks to advance the argument that only through improvements in the capacity of humans to manage technology, can human agency survive into a technological future. This argument therefore draws from Tegmark’s (2017) logic, that given uncertainty about technological outcomes, an important goal is to immediately undertake technology-safety research, and make this research mainstream. As an organising framework, literature is used to derive six primary technological
threats. These threats are used to guide and anchor discussions and arguments throughout different sections of the paper.

This paper therefore seeks to make a contribution to the future studies literature, in the following ways. First, it seeks to relate the threat of technological development to Sardar’s (2010) four laws of future studies. Whereas most approaches to such problems suffer from the constraints associated with disciplinary lenses, this work seeks to understand these problems through a systematic approach that does not privilege any specific disciplinary approach over another. Future studies “is not just multi- and trans-disciplinary, it is unashamedly un-disciplinary: that is, it consciously rejects the status and state of a discipline while being a fully fledged systematic mode of critical enquiry” (Sardar, 2010:183). This is Sardar’s first law, and it is particularly important in light of certain of the arguments made here, which draw on Nielsen’s (2012) theory of networked science, and its predictions that human collaboration enabled by novel technologies can result in radical innovations in the scientific discovery process itself. Nielsen argues that it is changes in the scientific discovery process itself that account for the great scientific leaps, or scientific advances through history. It is argued here that Nielsen’s theory offers useful insights into how radical innovations in the research, and R&D process itself can improve the capacity of humans to manage technological advancement, or other ‘wicked problems’ associated with technological proliferation, considered here in terms of six primary technological threats facing humankind.

Second, although seemingly alarmist, the potential threats of technological proliferation require consideration for those seeking to develop useful future scenarios for how technology may impact society going forward. Technological advancement has seemingly considerable potential for technological development and proliferation to contribute to the betterment of society through elevating the health and wellbeing of populations (and through offering outright solutions to catastrophic problems). However, it also has the potential for catastrophic consequences of technological development itself and the proliferation of harmful outcomes. Effective theory needs to be developed that can help stakeholders capture benefits while mitigating risk. This paper therefore poses an important question, namely what are the key theoretical dimensions of this problem? This problem is defined as the need to manage the seemingly unlimited opportunities of technological change while at the same time also managing the proliferation of dangerous technological applications.

In light of this problem, and the lack of sufficient knowledge of how to address the tensions between potential benefits and risk (and the failure of humankind to solve many of its present contemporary problems), this paper seeks to present an argument that two aspects of technological change account for most of the variance associated with this problem. It is also argued that a focus on the management of these two aspects, or dimensions, can provide important guidelines for the development of a discovery system that is relatively more robust to threats posed by both innovation failure as well as from dangerous technology proliferation.

These two dimensions relate to the urgent need (i) to manage the openness of the discovery system itself, and (ii) to manage the power relationships associated with a rapidly developing system of technological innovation. Sardar’s (2010:183) second law states that future studies are characterized by Mutually Assured Diversity (MAD), whereby MAD is “is the proposition that full preservation of our humanity requires that [human diversity] is assured, that it not only survives but thrives in any desired future, and that future generations mutually recognise and appreciate each others’ diversity.” It is argued here that key to the survival of humans in the face of technological proliferation, following Nielsen’s (2012) theory, is the need for openness in the discovery process itself, whereby openness is a necessary condition for the knowledge creation necessary for the co-evolution of our capabilities to manage the accelerating rate of emergent technologies. Indeed, according to Tegmark (2017:89), the challenges of technological advancement “transcend all traditional boundaries- both between specialties and between nations.”

Sardar’s (2010:183) third law relates to how futures studies are sceptical, and “the reality of things is inaccessible to the human mind and certitude is impossible to attain.” Given the uncertainties associated with technological proliferation, in developing theoretical insights it is necessary to take recourse to theory that predicts how diversity in human action has seemingly been checked by power throughout human history (see Foucault, 1982). Thus, to develop a theoretical frame that is sceptical of truth claims, the initial structure developed here focuses first on a key defining issue of the knowledge age, namely the radical changes in the discovery process on account of the democratisation of knowledge (Callaghan, 2016). These changes seem to be fundamentally enabled by the increasing openness of the knowledge creation process itself. Second, the analysis focuses on the way that power relationships have seemed to pose a fundamental threat to the diversity of views in human history (Foucault, 1982). In this way, analysis is related to primary technological threats.

Third, in light of Sardar’s (2010:184) fourth law, that future studies are futureless, and the influence of futures studies is on thought and scientific behaviour in the present, this paper seeks to “change peoples’ perceptions, make them aware of dangers and opportunities ahead” and to “galvanise them into collective social action” to provoke discussion and consideration of important and uncomfortable scenarios that will unfold on account of our actions in the present.

Having introduced the paper, and having provided a brief overview of its key arguments, literature is now reviewed, and certain key technology threats are identified. After this, interrelationships between these threats are discussed. A theoretical framework is then developed, and four modes of discovery are derived from the use of two propositions as heuristics. Certain issues drawn from the literature are then discussed in relation to each of these modes of discovery. The paper concludes with a discussion of these modes.

2. Theory and technological threats

The industrial and digital (information) revolution has had far-reaching impacts on “practically all aspects of our society, life, firms and employment” (Makridakis, 2017:46). As stressed previously, these effects are expected to intensify over time. For Makridakis (2017:50), those predicting different influences of technology on society can be differentiated on the basis of their optimism or pessimism:
[O]ptimists predict a “science fiction,” utopian future with Genetics, Nanotechnology and Robotics (GNR) revolutionizing everything, allowing humans to harness the speed, memory capacities and knowledge sharing ability of computers and our brain being directly connected to the cloud. Genetics would enable changing our genes to avoid disease and slow down, or even reverse aging, thus extending our life span considerably and perhaps eventually achieving immortality. Nanotechnology, using 3D printers, would enable us to create virtually any physical product from information and inexpensive materials bringing an unlimited creation of wealth. Finally robots would be doing all the actual work, leaving humans with the choice of spending their time performing activities of their choice or working, when they want, at jobs that interest them.

Pessimists, however, offer a more dystopian view while others are more nuanced in the way they frame opportunities or threats associated with technological advancement. With regard to AI, Tegmark (2017) differentiates between different schools of thought on the basis of when they consider AI to ultimately be able to surpass human levels of intelligence, and whether they consider this ‘superhuman’ AI to be a threat or not. This event has been termed the ‘singularity.’

2.1. The control threat

According to Vinge (1993:1), if the technological singularity “can happen, it will,” and we should therefore primarily be concerned about issues of control. The challenge of losing control of technological advancement is termed the ‘control threat.’ Long-standing debates in the literature seem to suggest that this control threat warrants consideration as a primary technological threat. Vinge (1993:1) defines the singularity as the “imminent creation by technology of entities with greater than human intelligence”.

Bostrom (2017:300) refers to the singularity as an “intelligence explosion,” whereby “we humans are like small children playing with a bomb,” such “is the mismatch between the power of our plaything and the immaturity of our conduct.” Here, issues of human agency are central to the dangers inherent in this loss of control. Loss of control of dangerous technologies, and the need for responsible innovation (Grunwald, 2011; Stilgoe et al., 2013) are therefore key to discussions of the control threat. For Bostrom (2017:300), in light of the threat of an intelligence explosion the “most appropriate attitude may be a bitter determination to be as competent as we can, much as if we were preparing for a difficult exam that will either realize our dreams or obliterate them.”

In light of increasing uncertainty related to technological advancement, it is necessary to consider all voices, however abhorrent, or diversity of opinion, in “all of its mindboggling forms” (Sardar, 2010:183). According to Kaczynski (1996:1), as “society and the problems that face it become more and more complex and machines become more and more intelligent, people will let machines make more of their decisions for them, simply because machine-made decisions will bring better results than man-made ones.” Kaczynski (1996) argues that the increasing complexity associated with decision making will then ultimately outstrip the capacity of humans to make them, and human control of decisions would be lost.

The control threat therefore relates to the problem that humankind will simply not be able to manage these changes, primarily due to increasing complexity. Tegmark (2017) argues that many in the scientific community subscribe to the beneficial AI movement, whereby survival of a technological future depends on threats being identified now, and safety research so as to ensure that technological development will be beneficial to humankind. Central to this argument is the notion that the competence of the management of technological advancement will determine how beneficial is this advancement. Accordingly, binary conceptions of technological advancement as either utopian or dystopian might yield unhelpful and erroneous heuristics.

Tegmark (2017) stresses that discussions of malevolent machines are a red herring, and that the control issue reduces to one of competence, whereby the true danger would exist if the goals of a competent AI, for example, diverged from ours. It is argued here that technological advances need to be harnessed to enable human collaborative problem solving, or the capabilities associated with human collective intelligence so as to address this control threat, and to ensure that human management of technology is not eclipsed by machine intelligence or outpaced simply by the increasing complexity of the management challenge itself.

2.2. The power inequality threat

There are also challenges that might arise due to this loss of control. Certain of these challenges relate to how power can influence technological change, and in turn societies, and particularly how elites might behave if their power is unchecked. The loss of control of dangerous technologies to powerful elites, be they business, political or otherwise, can also have important consequences, just as it would if control is lost to researchers, absent principles of responsible innovation. Having the management of dangerous complex technologies subject to market profitability logics can be problematic, as shown in research on nuclear technology (Osborne & Jackson, 1988). Losing control of dangerous technologies to market power, or exposing them to the vagaries of executive risk seeking are therefore yet other dimension of the control threat.

The question naturally arises as to who should then be in control of technological advancement? According to Olsen, Kruke, and Hovden, (2007:69), societal safety is “a sensitive political issue containing dilemmas and value choices that are hardly possible to perceive or solve as pure scientific problems.” Central to such a perspective is the role of power in how such issues come to be managed. Given multiple perspectives and stakeholder interests, the full reality of things as they relate to dangerous technologies may indeed be impossible to perceive accurately (Sardar, 2010:183). The dispersion of power, including that associated with knowledge, may ultimately be key to, at the very least, ensuring societal scrutiny of the issues associated with dangerous technologies.

Power is of central importance in understanding the consequences of human behaviour (Foucault, 1982). There are therefore also perhaps increased dangers posed by elites if human control is maintained, and if technological change increases this power.
According to one dystopian perspective, human work may ultimately no longer necessary, and “the masses will be superfluous, a useless burden on the system” (Kaczynski, 1996:1). The following passage is from Kaczynski (1996:1):

If the elite is ruthless they may simply decide to exterminate the mass of humanity. If they are humane they may use propaganda or other psychological or biological techniques to reduce the birth rate until the mass of humanity becomes extinct, leaving the world to the elite. Or, if the elite consists of soft-hearted liberals, they may decide to play the role of good shepherds to the rest of the human race. They will see to it that everyone’s physical needs are satisfied…Of course, life will be so purposeless that people will have to be biologically or psychologically engineered either to remove their need for power processes or make them “sublimate” their drive for power into some harmless hobby. These engineered human beings may be happy in such a society, but they will most certainly not be free. The will have been reduced to the status of domestic animals.

The notion of the need to biologically or psychologically engineer away the need for power differs radically from Foucault’s (1982) approach. To consider future technological scenarios, however, one cannot shy away from unpleasant narratives, no matter how ‘wicked’ or seemingly impossible to solve (Sardar, 2010), and this is a case in point. Joy (2000) also cites this paragraph, acknowledging Kaczynski’s role as a contemporary Luddite, and stressing the need to confront the arguments made therein. The problem of power inequality that may result from technological change is termed here the power inequality threat. Technologies such as human reproductive cloning and inheritable genome modification (Bostrom, 2017; Isasi & Knoppers, 2015) also raise important issues about inequality that may result of genetic engineering, including concerns about eugenics (Gilding, 2002).

There is clearly a need to address these issues, but in a different way than Kaczynski, who as the Unabomber sought to physically attack (bomb) those involved in the advancement of technology, including those in universities (Joy, 2000). Indeed, the power inequality threat might not extend simply to elites, and the power to unleash destructive forces would perhaps give those (ironically, such as Kaczynski himself) who seek destruction (for any reason) the means to do so, and ultimately, perhaps, the ability to threaten humankind itself.

2.3. The destructive empowerment threat

Advances in technology can alter the balance of power between nations and between individuals and security institutions. Tegmark (2017:107) stresses that “those who stand to gain most from an arms race aren’t superpowers but small rogue states and non-state actors such as terrorists, who gain access to the black market once they’ve been developed.” Cyberwar, and its potential to disable critical infrastructure, will become increasingly likely between belligerent states as technology advances. Joy (2000:1) stresses that genetic engineering “gives the power- whether militarily, accidentally, or in a deliberate terrorist act- to create a White Plague.” According to Joy (2000:1), nanotechnology has “clear military and terrorist uses, and you need not be suicidal to release a massively destructive nanotechnological device- such devices can be built to be selectively destructive, affecting, for example, only a certain geographical area or a group of people who are genetically distinct.” Further, replicating nanotechnology can be dangerous on its own, for example in the way they can crowd out other life, and this could result from a single laboratory accident (Joy, 2000).

Given a world population of billions of individuals, under the assumption that there are very few at the far end of a normal distribution of harmful intentions, there would perhaps always be some that would meet Joy’s definition of ‘extreme individuals.’ With the proliferation of increasingly dangerous technologies, and increasing access to them, these scenarios are perhaps increasingly likely. This threat is considered here the destructive empowerment threat. Another example offered by Kurzweil (1999) relates to the dangers of nanotechnology, which might be more dangerous than nuclear, as its consequences are not localised, but can spread. According to Kurzweil (1999:160), once “the basic technology is available, it would not be difficult to adapt it as an instrument of war or terrorism.”

2.4. Intrinsic challenge displacement threat

Issues related to the loss of purpose, such as those suggested by Kaczynski (1996), that might be experienced by those who’s need to work has been displaced by more effective and efficient technologies is considered the intrinsic challenge displacement threat. Indeed, the psychological impact of such technological change is uncertain, and consideration of this issue seems warranted. For example, Tegmark (2017:89) stresses the need to “grow our prosperity without leaving people lacking income or purpose.” This threat is also taken to relate to how technology might replace humans in the workplace, for example in ways that result in ‘technological unemployment amidst digital plenty.’ Brynjolfsson refers to this future scenario as a ‘digital Athens,’ whereby in the same way that Athenian citizens lived lives of leisure with labour performed by captives, a highly productive and automated economy might free up human labour without reducing living standards (in Regalado, 2012).

The reality, however, may be very different. Some have argued that we are entering an era of machine intelligence that ultimately heralds the end of human employability (Brynjolfsson and McAfee, 2011:9). Rifkin (2011) argues that the world is experiencing a third industrial revolution related to computer power (following the first, associated with steam power, and the second, related to the rise of oil and electricity as sources of power). According to Rifkin, machines are increasingly displacing human jobs and making blue collar work obsolete, giving rise to a ‘silicon-collar’ workforce, of machines that have replaced humans in the workplace. An ‘end of work’ era of worker economic irrelevance, and extensive joblessness might give rise to problems like rising levels of crime and feelings of irrelevance and alienation (Rifkin, 1995). Closely related to this threat is that of resource competition, in the form of competition from machines for human jobs.
2.5. The resource competition threat

The resource competition threat can become an unintended societal consequence of technological advancement. Joy (2000) points to how the design and use of technology has resulted in unintended consequences, such as the overuse of antibiotics which has given rise to antibiotic resistance, and drug resistant genes in Malaria parasites. Drawing from Moravec’s (1999) work, he also points to an argument that biological species “almost never survive encounters with superior competitors,” suggesting further that in a “completely free marketplace, superior robots” would outcompete humans, as robotic industries “would compete vigorously among themselves for matter, energy, and space, incidentally driving their price below human reach,” whereby humans, unable to afford the necessities of life, could be “squeezed out of existence” (Joy, 2000:1). This type of problem relates to the problem of crowding out, primarily related to resources, or the resource competition threat.

Brynjolfsson and McAfee (2014) point to how technological change may ultimately displace jobs. Advances in technologies are creating an unprecedented reallocation of wealth and income. However, whereas wages have increased alongside productivity for the previous two centuries, median wages have recently stopped tracking productivity, with important societal implications. Brynjolfsson and McAfee (2014) stress that on account of these changes, technological change has altered certain structural economic-relationships and that this is in turn driving rapidly-increasing inequality. A small elite are therefore benefitting from growth in GDP and productivity but the median income is diverging from the mean. The main driver of this increasing inequality is therefore exponential, digital and combinatory technological change driven by the new economics associated with near zero marginal cost (Rifkin, 2014) which is creating ‘winner takes all’ markets where leading providers (with a fraction of traditional employment costs) can capture most of a market through digitization.

Brynjolfsson and McAfee (2014) argue that it is no longer true that a rising tide of technical progress will ‘lift all boats’ because technology acts as a multiplier, in that while it produces more with limited inputs, it also substitutes for workers in lower-skilled work, increasing returns to high skilled work. This then results in skill-based technical change associated with increasing inequality.

Technology is also shifting the returns to physical capital versus labour. The corporate profit share of GDP has now surpassed that of the wage share, bolstered by winner-take-all markets enabled by the low marginal costs of digital goods and their low capacity constraints which allow substantial economies of scale (Brynjolfsson & McAfee, 2014). Others however have argued that previous dystopian predictions of devastating technological unemployment have in every historical case failed to materialise. According to Tegmark (2017) it may be different this time, as those arguing this might not have considered what will occur when machine intelligence begins to perform the creative work at which humans currently outperform machines.

2.6. The reproduction threat

According to Joy (2000:1), robots, engineered organisms, and nanobots differ from all previous technologies, as they share the ability to self-replicate, and with this will necessarily come the risk of substantial damage to the physical world; with each of these technologies, a “sequence of small, individually sensible advances leads to an accumulation of great power and, concomitantly, great danger.” There is therefore an issue related to the power of humans versus machines, and the dimensions along which these differences in power can result in different scenarios for society. Although this also derives from the control threat, this threat is considered more specifically to relate to the reproduction of technology, and the potential for exponential increases in the harmful effects of technology. This threat is defined here as the reproduction threat. The unchecked reproduction of nanomachines is a particular concern, according to Kurzweil (1999:158), due to the fact that to “be effective, nanometer-sized machines need to come in the trillions” and the “only way to achieve this economically is through combinatory explosion: let the machines build themselves.” He points to the risk of an exponentially exploding nanomachine population, and the risk of even minor software problems that fail to halt self-replication. The same is true for other technologies such as biotechnology. According to Kurzweil (1999:176), we are “very close to the point where the knowledge and equipment in a typical graduate-school biotechnology program will be sufficient to create self-replicating pathogens.” A theoretical model that seeks to contribute useful insights regarding the management of these threats therefore also needs to take into account the amount of time available to do this.

2.7. Scenario timelines

How long do we have until we can no longer manage technological advancement and its proliferation? According to Tegmark (2017), two conferences of AI researchers have collectively estimated that human-level artificial general intelligence will be created by the year 2055 (the first conference) and 2047 (the second, two years later). With regard to when dangerous technologies may no longer be manageable, Joy’s perspectives are included here as being representative of many dystopian commentaries. Unlike the 20th century weapons of mass destruction, GNR technologies are being rapidly developed commercially by corporate enterprises, and their promises are being aggressively pursued; according to Joy (2000:1), this “is the first moment in the history of our planet when any species, by its own voluntary actions, has become a danger to itself- as well as to vast numbers of others.”

Although it is simply not known when the machine intelligence ‘explosion’ will occur, other dangerous technologies are currently also developing; therefore, in terms of timelines, technology safety research is an immediate imperative according to Bostrom (2017) and Tegmark (2017). This is also the argument of Vinge (1993:1), whereby if the technological singularity “can happen, it will;” we should therefore primarily be concerned about issues of researching how this can be prepared for. Our scientific research capacity is therefore key to our ability to undertake effective technology safety research, and to our ability to manage dangerous technologies so as to not lose control of them.
Recent literature suggests that timelines in scientific research production may be about to shorten. In terms of the research capacity needed to support technology safety research, there is another perspective that argues that the human capability to research, and therefore to manage, a threatening context is about to be radically enhanced. Whereas the literature to date seems to have given much attention to scenarios similar to Joy’s (2000) commentary, lacking in these debates seems to be the argument that science itself is on the cusp of a reorganization, which some have termed the ‘reinvention of discovery.’ According to Tegmark (2017:169) the “reinvention of discovery is one of the great changes of our time” whereby to “historians looking back a hundred years from now, there will be two eras of science: pre-networked science, and network science” as we are currently “experiencing a time of transition to the second era of science.” The theoretical model to be introduced in the following sections will draw on this body of literature [networked science] that suggests the emergence of a new dawning of knowledge creation, whereby novel technological developments make it possible to take advantage of hitherto unimagined economies of scale in both (big) data collection and, importantly, analysis.

If the threats posed by technology are real, and if relinquishing science (as advocated by Joy) will not solve the problem of technological proliferation, and only exacerbate the current inequalities in innovation outcomes, then there is perhaps only one other alternative, namely to enhance our ability to manage it. According to networked science theory, this might be possible. For the purposes of this work, the opposite of relinquishment is taken to be uptake of open modes of science, or open innovation. Other alternatives arguably fall into these two categories, or along a continuum between the two. At this nexus, interrelationships between the technological threats discussed above are discussed, and then ‘real life’ examples are considered that specifically relate to the complexities associated with the management of dangerous technologies.

3. Interrelationships between the technological threats

To develop a theoretical model of technological threats and their potential impact on society it is first necessary to relate these threats and to derive underlying regularities between them as the basis for a problem solving response that can to some extent address aspects of them all. Before discussing this model it is then necessary to consider existing proposed solutions to the threats discussed above.

The only realistic alternative [to the dangers of technological advancement], according to Joy (2000:1) is “relinquishment: to limit the development of the technologies that are too dangerous, by limiting our pursuit of certain kinds of knowledge.” Tegmark (2017:169) acknowledges that although the term Luddite is typically used as a derogatory epithet for those who are perceived as technophobes “on the wrong side of history,” the notion of relinquishment has nonetheless found “new support” in the environmental and anti-globalisation movements. This is an important argument, because to revalue knowledge as either ‘bad’ or ‘good’ would seemingly have important implications for almost every aspect of human life. Whereas six broad threats were derived in the previous sections, the notion that knowledge can be harmful is an important proposition, and this idea (relinquishment) as a proposed solution requires interrogation according to Joy’s criteria of potential harm, using these same six categories of threat as heuristics to sharpen the discussion.

The nature of innovation itself, and how it works through different channels, is not independent of these issues, and its channel of open innovation presents exponential advantages for knowledge creation, whether from its ability to harness very large volumes of data, or to harness large scale data analysis, or problem solving opportunities (Callaghan, 2016). This potential has been described in terms of the contributions of collective intelligence (Bernstein, Klein, & Malone, 2012) and networked science (Nielsen, 2012), which share a focus on the opportunities offered by open science, and open systems of innovation.

In contrast, Joy (2000:1) argues further that “despite the strong historical precedents, if open access to and unlimited development of knowledge henceforth puts us all in danger of extinction, then common sense demands that we reexamine even these basic, long-held beliefs.” One therefore has to question how realistic Joy’s solution (relinquishment) really is, to what is clearly a wicked problem according to Sardar’s (2010) futures study approach. There are arguably three potential problems with the relinquishment approach.

Firstly, if open access gives way to closed access, and some (not all) relinquish knowledge, but not others, then who would access to this knowledge be limited to? The power inequality threat is not addressed by shutting down open access. Control of technology might then shift to elites, and as we surely know by now, state powers to limit or suppress activity can be captured by powerful interests, essentially creating the conditions for a monopoly in knowledge. Thus, secondly, the control threat brings with it similar problems as those associated with losing control of technology to intelligent machines. Thirdly, there is the problem posed by slow response to threats under closed systems of innovation (discovery). Under closed systems, destructive empowerment threats would perhaps be more difficult to defend against, without the social and institutional infrastructure that open systems of innovation are rapidly developing. Open systems of knowledge creation have demonstrated their increasing effectiveness in enabling timely disaster response (Callaghan, 2016). To contextualise certain of these ideas prior to discussions of the theoretical model it is necessary to first identify certain real life examples of the tension between openness and closure in decisions about dangerous technologies.

3.1. Contextualising the reality of technological threats

Certain decision making issues related to dangerous technologies have already been considered in real world contexts. Dual-use research of concern (DURC) is perhaps a useful example of how issues related to the control, power inequality, and destructive empowerment threats have been considered to date. DURC research is “research that, based on current understanding, can be reasonably anticipated to provide knowledge, products, or technologies that could be directly misapplied by others to pose a threat to public health and safety, agricultural crops and other plants, animals, the environment, material, or national security” (NIH, 2017:1).
An example of DURC research can be found in the debates concerning the publication of two papers revealing how to genetically engineer strains of the H5N1 avian influenza virus (Resnik, 2013). Those arguing against publication have cited concerns, particularly since 2001, about the use of this knowledge by terrorists, or others with destructive motives (Resnik, 2013), to create a bioweapon and set loose a global pandemic (Cohen & Malakoff, 2012). According to Specter (2012:1), the decision to allow publication of this knowledge “fundamentally altered the scope of the biological sciences.” The U.S. National Institutes of Health in 2011 initially recommended redaction of these papers, but after careful consideration the National Science Advisory Board for Biosecurity (NSABB) recommended (notwithstanding a lack of unanimity) they “should be made public, in full,” as the potential public health benefits were considered to outweigh the potential harm (Cohen and Malakoff, 2012:19). This decision (creating an important precedent) was therefore taken in support of openness rather than closure, notwithstanding the destructive empowerment threat.

Over and above the issue of freely available dangerous information, accidental release of pathogens is another threat that is known to occur, if not regularly, but then often enough to warrant consideration here. Evidence-based examples of fatalities exist in the form of research-related accidental release of smallpox, severe acute respiratory syndrome (SARS), and Ebola pathogens (Specter, 2012:1). Prior to the NSABB decision, in 2002, someone had already “stitched together hundreds of DNA fragments, mostly acquired via the Internet, then used them to create a fully functional polio virus,” and in 2005 academic papers published the genomic sequence of the 1918 Spanish flu, but these have both (notwithstanding much initial condemnation) ultimately been considered valuable contributions to knowledge (Specter, 2012:1).

The threat of biological terror seems real, as even Al Qaeda have called for its supporters with degrees in microbiology or chemistry to develop a weapon of mass destruction (Specter, 2012). This threat is of great concern, given proof of concept of how relatively easy it has been to reconstitute an extinct poxvirus, costing approximately $100 000 using only mail-order DNA (see Kupferschmidt, 2017).

Some also argue that such knowledge can help those developing vaccines or drugs to know if these are effective. Additionally, the scientific method and “the entire edifice of institutional research depends on such openness; without it, progress would slow dramatically” (Specter, 2012:1).

However, unlike the all-or-nothing decisions to research or produce pandemic strains of pathogens, the threats of GNR technologies are unclear, even as they currently proliferate, mostly behind the closed doors of commercial enterprises. At the same time, the artificial intelligence (AI) revolution will bring extensive changes to all aspects of society and life, and additionally to firms and employment, “resulting in richly interconnected organizations with decision making based on the analysis and exploitation of ‘big’ data and intensified, global competition among firms” (Makridakis, 2017:46). Previous research has sought to make sense of the complexity of human engagement with technology through the use of metaphors to describe technological futures. This literature is now also briefly considered here to contextualise discussions in the above sections.

3.2. Metaphors as a heuristic for understanding technological threats

The evolution of technology and its core threats can be taken to be reflected in the metaphors people use when considering technological futures. Metaphors used by stakeholders reflect the evolution of technologies, as for the past two centuries the ‘technology is good’ metaphor has persisted, related to improvements in productivity; this metaphor has also been associated with another, namely that ‘more is good’ (Carbonell, Sánchez-Esguevillas, & Carro, 2016). Joy’s (2000) perspective might be read as a metaphor, that ‘technology is dangerous,’ conflicting with the metaphor that ‘technology will solve our problems.’ Drawing directly from this is the binary conflict between the metaphors ‘technology should be relinquished,’ and therefore that ‘closed models of development are best’ versus ‘technology should be shared, and open models are best to keep us safe.’ However, the danger here is that subscribing to these metaphors simply puts us at risk of creating unhelpful binaries.

It goes without saying that there are always gradations between these extremes, and Carbonell et al.’s (2016) use of technology metaphors are useful in order to simplify explanations of conflicts between different perspectives. These metaphors are then useful as heuristics, in that they can be related to the six technology threats, encouraging dialectical tensions that give rise to a more considered discussion of scenarios. On the basis of these conceptions, the answering metaphor is perhaps that ‘technological dangers can be successfully managed,’ juxtaposed against its counterpoint ‘technological dangers cannot be successfully managed.’ The relinquishment argument of Joy (2000) might needlessly echo historical Luddite arguments if there are no other options with which to frame our response to technological dangers. We might have no other choice but to embrace open systems of discovery in order to improve our ability to manage technology, with the hope that improved systems of discovery will ultimately be key to the successful management of technological change itself.

Historical Luddite protests associated with the metaphor ‘the job is up’ rather than ‘technology is up’ offer an early example of debates about the trade-offs some argue are to be made when technology advances (Carbonell et al., 2016). This is perhaps an example of the resource competition problem and the threat posed by technology to resources in the form of jobs.

Other examples include those of religious groups that have also advanced metaphors conflicting with technology, as reflected in longstanding historical tensions between science and religion. Similarly, there are now tensions between societal values like equality, respect or privacy (reflected in concerns about the digital divide, harassment and other outcomes) and the capacities the Internet now offers (Carbonell et al., 2016). These tensions can perhaps be related to the control threat, as individuals face losing control over privacy, over the continuity of their lifestyles, or as societies lose control over widening inequality on account of the digital divide. The latter issue also relates to the power inequality threat. If the cat is out of the bag already (as the example of the publication of the H5N1 papers shows), and relinquishment may no longer be a useful strategy (as countries and individuals differ in their moral propensity to develop and use dangerous technologies (Boström, 2017)), the only way out may be to radically increase our ability, as
humans, to collectively manage these threats.

As it stands, it is unlikely however that we have this capacity at present or will be able to develop it quickly. How then, could this capacity be developed? And what future scenarios would result from failure to successfully manage these challenges? Alternatively, what future scenarios would result if such successful management of technological proliferation were possible? Successful management is defined here as effective research and knowledge creation that enables the threats of technological development to be mitigated in a sustainable way (Bostrom, 2017; Tegmark, 2017), and which results in a relatively equitable distribution of the outcomes of discovery (Rifkin, 1995).

One would need to ask, however, what is the role of the state in such successful management? Other metaphors relevant to debates about the impact of technology on society are those related to the tensions between ‘big brother dystopia’ versus ‘state as protector,’ and ‘equality is up’ versus ‘market is up’ (Carbonell et al., 2016). The enhanced surveillance abilities of the state might also lend themselves to a change in power dynamics, and an increase in power inequality, as this power might be part of the trade-off for safety in an era in which public gatherings, for example, are increasingly vulnerable to attack. Indeed, the same technological advances can also enable destructive empowerment, as individuals can use technology to amplify the damage they can cause. The key then, might be to therefore consider such management according to the principles of openness, and the democratisation of science, and its attendant ethical framework. According to the principles of maximum transparency and accountability, power inequality is reduced, and power over knowledge is made to be more equitable. In this way, inequality in the outcomes of knowledge is also reduced, maximising benefits to all affected by science as well as the problems it is tasked to solve.

A key feature of the theoretical model proposed here to offer certain insights into the societal impact of technology is therefore open knowledge creation, and an ethical framework that is fundamentally suited to open systems of innovation and discovery. Given scarce resources (including time), however, it is unclear as to which of the six threats require more urgent attention than others. What then are the relationships between these threats?

3.3. Ordering relationships between threats

Fig. 1 shows a possible ordering of technological threats. As discussed above, these threats reflect primary technological concerns in the technology futures literature.

Criteria for inclusion was based on the perceived relative seriousness of a threat. Threats were not considered for discussion on their own if they fell within another of these categories, other than the control threat category. This iterative and inductive process of review resulted in the six categories included here. A brief sketch is now provided, of how these threats might relate to each other.

![Fig. 1. Interrelationships between the six technological threats.](image-url)
Technological futures are by definition uncertain, and the relationships discussed here are necessarily speculative, but such a discussion is necessary in order to draw out an ordering of these threats and to better understand which are more urgent.

If the control threat is considered the ‘origin’ of the other threats considered here, then the management of this threat would require an immediate and proactive response. This threat is therefore ‘immediately urgent’ while those that derive from it are ‘urgent,’ in that the control threat would need to be considered together with the others.

Such an ordering might have important implications for which societal stakeholders should be more involved in managing these threats. If the control threat is considered the dominant threat, then this places technology safety researchers at the source of the problem of managing dangerous technologies. Indeed, if technological development to date has typically followed the trial-and-error model, then we will “inevitably reach the point where even a single accident could be devastating enough to outweigh all benefits” (Tegmark, 2017:90). Having private or corporate stakeholders drive the technological development process without the engagement of independent research stakeholder groups may no longer be safe in an era that transcends trial and error approaches to dangerous technologies.

According to the logics described in Fig. 1, the control threat, or losing control of the management of technology, can therefore lead to other threats. We are now perhaps in an era in which the consequences of practice-based trial and error make it necessary to elevate the status of technology safety researchers. This is an immediately urgent imperative.

Certain research findings seem to support the necessity of a change in societal stakeholder relationships (as they relate to dangerous technologies) to include technology safety researchers. Insights from the use of nuclear power suggest that certain risks can arise from organisational structures of large corporate organizations themselves. Fewer coordinative mechanisms between functional departments, more levels of administration, centralization and higher numbers of employees may constrain an organization’s ability to respond to safety issues (Osborne & Jackson, 1988). Risk preferences are also not constant over different types of decision making (Osborne & Jackson, 1988). Indeed, under conditions of growing losses, decisions are typically more risky than they are under conditions of gains (Kahneman & Tversky, 1979). Osborne and Jackson (1988:930) therefore suggest that the proportion of a utility’s investment in a dangerous technology like nuclear power “partially reflects the technological risk preferences of its senior executives.” Developing and empowering technology safety researchers as an important stakeholder group may therefore be an urgent need, so as to ensure that the post-trial-and-error paradigm is safely managed through more inclusive engagement going forward.

To manage the control threat it may be important to therefore shift the locus of power related to decision making about dangerous technologies from corporate or other interests to proactively include technology safety researchers. This dispersion of power might however be at odds with historical practice and the autarky of corporate R&D. The management of, and decisions about, dangerous technologies require openness and societal inclusion, according to the principles of responsible innovation (Grunwald, 2011; Stilgoe et al., 2013). According to Douglas (2000:559), “value-free science is inadequate science; the reasoning is flawed and incomplete.” The task of managing the control problem cannot therefore simply be left to corporate market incentives, or even to science on its own.

Thus, the threat of losing control of technology, whether to human elites or to machine intelligence, is perhaps the most important of the threats, and is considered immediately urgent, requiring inclusive engagement across society. This implies openness and power dispersion to mitigate against loss of control of dangerous technology to any set of particular interest groups. Given the centrality of the control threat, what then of the relationships between the others described in Fig. 1?

The control threat, if unsuccessfully managed, may contribute to the power inequality, resource competition and destructive empowerment threats. The resource competition threat may in turn contribute to increased power inequality through two channels. The first is arguably the way digitisation is creating a winner-takes-all economy (Brynjolfsson & McAfee, 2014) as the new economics of near zero marginal costs (Rifkin, 2014) allow a few producers to capture substantial market shares with very few human workers. Another channel might be through the erosion of jobs that fall below the waterline (Tegmark, 2017). If technology creates or exacerbates such class differentials, and if these classes are able to prioritize their own interests at the expense of others, they might seek solutions associated with power inequality.

Intrinsic displacement might be considered a derive threat, arising from a lack of purpose in a world in which machines do most forms of work, or (alternatively) a state of powerlessness as people are excluded from meaningful opportunities by a technologically-enabled human elite. Thus, the four threats of control, resource competition, power inequality and intrinsic displacement might benefit from further research that considers their potential interdependencies.

The destructive empowerment threat is an ever-present one, as advancing technology necessarily provides more options for both individuals and states to pursue destructive goals. Although the link is not shown in Fig. 1, power inequality can result in destructive empowerment if elites or elite countries use these technologies in war. A global power hierarchy held in place by technology might be such an outcome. Key to managing the relationships between these threats, however, seems to be the need for proactive engagement with technology safety researchers, and the use of technology to improve our research capabilities and, thereby, our ability to manage technological change.

Thus an order of importance seems to exist amongst these threats. A focus of resources and attention on the control threat without neglecting relationships between these threats is important. This approach seems to also find support in the literature. Bostrom (2017) and Tegmark (2017) stress the urgency of technological safety research, to be able to control the trajectory of technological change, and ensure its ‘beneficial’ use and contribution to human society. Their arguments capture the essence of what the management of the control problem entails.

If managing the control threat is key to the management of the others, what then are the principles most likely to empower this control, or management of technology? The theoretical model that follows draws from novel ideas and theory that suggest certain
principles that might be useful in this task.

4. Theoretical model

According to Tarko and Aligica (2011:987), Kahn’s conceptualisation of the ‘institutionalisation of interdisciplinarity’ is reflected in “a phenomenon that has as its core a process-based approach to knowledge and method aggregation” reflected in novel developments enabled by web-based techniques. According to Tarko and Aligica (2011), terms associated with this phenomenon include Wikinomics (Tapscott & Williams, 2006), the wisdom of crowds (Surowiecki, 2004), the “army of Davids” (Reynolds, 2006), and “collective intelligence” (Malone, Laubacher, & Dellarocas, 2009). Malone et al. (2009:2) also acknowledge other additional descriptions of the phenomenon in the literature, such as radical decentralization, crowd-sourcing, and peer production, arguing that the term collective intelligence is the most useful, defined broadly as “groups of individuals doing things collectively that seem intelligent.” The phenomenon has also been defined as crowdsourced R&D, and considered in terms of its roots in the seminal knowledge aggregation problem with a view to formalising this both as a body of theory and as a new scientific methodology in its own right (Callaghan, 2016). Nielsen (2012) argues that dramatic breakthrough periods in science have typically followed changes, or improvements in the way discovery is conducted. For Nielsen (2012:19):

This change is important. Improving the way science is done means speeding up the rate of all scientific discovery. It means speeding up things such as curing cancer, solving the climate change problem, launching humanity permanently into space. It means fundamental insights into the human condition, into how the universe works and what it is made of. It means discoveries we’ve not dreamt of. Over the next few years we have an astonishing opportunity to change and improve the way science is done.

Rapid acceleration of the pace of scientific discovery, however, requires an ethical framework that is robust to the range of different issues that can be encountered. Over time there have been increasing calls for increased democratization of science, and for greater stakeholder involvement (Siune et al., 2009). Concerns about the role of science in society and its impacts have contributed to the rise of new research fields. These fields include risk studies, impact studies, technology assessment, [Science and Technology Studies (STS)], and applied ethics, which are increasingly integrated into research programmes (Siune et al., 2009). Governance of science and R&D processes is changing, opening up “new possibilities and opportunities for involving new actors and new types of reflector” (Grunwald, 2011:9).

This literature highlights a growing movement advocating the democratization of science premised on open models of knowledge creation. The democratization of science movement stresses the increasing importance of disclosure and transparency issues not only in the contemporary bioethics field, but in broader areas of technology governance. The concept of ethical practice in this literature highlights the importance of increased transparency together with increased accountability to stakeholders. This perspective echoes the emergence of new movements, such as those associated with citizen science (CS) (Bonney et al., 2009), public participation in scientific research (PPSR) (Shirk et al., 2012), and participant-led biomedical research (PLR) (Vayena & Tasioulas, 2013), which all relate to increasing access of citizens, or populations to the research process itself. These movements are in turn related to post-normal science (Funtowicz & Ravetz, 1994), and its ethical framework premised on the need for maximized transparency and accountability. The need for the post-normal science ethical approach arose from the tensions between different perspectives of climate science, whereby only through maximized transparency could the necessary scrutiny of research findings result (Funtowicz & Ravetz, 1994). These bodies of theory extend stakeholder theory (Freeman, 1984) and may form the basis for a complementary model of ethics in science that is robust to technological change.

Synthesis and integration of this literature suggests certain core ideas. The first is that the growing literature on technology governance seems to be able to provide ethical frameworks that might be sufficiently robust to support a rapid acceleration of the pace of scientific discovery. However, key to this is the need for maximised transparency and accountability, and for the full inclusion of societal stakeholders in technology governance.

The second is that, as suggested by Nielsen (2012) and documented in his work on networked science, there seems to be a coming ‘revolution’ in science itself, whereby we are on the cusp of a ‘second era’ of science, in that the nature of the research process itself is changing. In fact, Nielsen’s (2012) theory is perhaps foreshadowed by prior examples of the same phenomenon in the futures literature (see Tarko & Aligica, 2011). These changes also seem to echo Särddar’s (2010) principles, in that networked science transcends disciplines (first law), incorporates maximised diversity and inclusivity (openness) across society (second law), thereby mitigating the uncertainties (third law) inherent in the interactions of human agency with technological change though the dispersion of power and the empowerment of the scientific citizen. Such changes in the processes of science, or scientific research itself may make it possible to develop the management capabilities to address the control threat. A synthesis of this literature suggests, however, that there are two necessary (but not sufficient) conditions to the successful management of the control threat, namely the need for openness as a primary mode of discovery, and the need for dispersion in the power relationships around the management of dangerous technologies. From this body of theory, the following proposition is derived.

Proposition 1. The successful management of technology is fundamentally related to openness as the primary mode of discovery.

The need for maximized transparency and accountability is therefore taken to necessarily be related to openness, or open access to knowledge and information as a necessary condition. There is a need however to articulate the tensions between the six problems, or potential technological consequences, as knowledge of the complex interrelationships between these threats is important, given that solving one problem might exacerbate another. It is therefore necessary to construct solutions that address a substantial aspect of these problems at the same time.
Given a framework that maximises accountability and transparency, ethical management of technological change may be possible. Under closed modes of discovery, relinquishment of technology might not occur, as those with more power would not have to disclose what they have not relinquished. The relinquishment option may therefore not be as effective as an open mode of discovery in addressing technological threats, as long as transparency and accountability is ensured in an open environment. However, because transparency and accountability, as well as the ethical framework related to this, is but a necessary condition, and not a sufficient condition for the effective management of technological change, it is argued that a further condition is necessary, namely the need for dispersed power relationships, whereby dominant elites do not gain control of the discovery process, resulting in inequitable access to it, and unequal access to its outcomes.

It is with regard to the need for openness and for dispersed power relationships in discovery that we then need to weigh up Joy’s (2000) other alternative, namely to give up the goals of perpetual economic growth as they may be inseparable from the dangers of technological growth. Joy (2000:1) suggests that material progress and the pursuit of the power of knowledge are problematic goals, arguing that “we must find alternative outlets for our creative forces, beyond the culture of perpetual economic growth; this growth has largely been a blessing for several hundred years, and we must now choose between the pursuit of unrestricted and undirected growth through science and technology and the clear accompanying dangers.” Indeed, openness, even with its attendant ethical framework, might not on its own be enough to address this threat, but it is arguably only through control that human incentives for progress can be subdued. A better solution therefore might not be the curtailment of material progress but the mitigation of the power of knowledge associated unequal concentration. Whereby openness ensures access to information and knowledge for affected populations, what is additionally needed is a mechanism to ensure the dispersion of power, or a mechanism to address power inequality, and to address threats of control. In the seminal words of Foucault (1982:780):

I would like to suggest another way to go further toward a new economy of power relations, a way which is more empirical, more directly related to our present situation, and which implies more relations between theory and practice. It consists of taking the forms of resistance against different forms of power as a starting point. To use another metaphor, it consists of using this resistance as a chemical catalyst so as to bring to light power relations, locate their position, and find out their point of application and the methods used.

In order to optimise the ability of collaborative human networks to manage rapidly developing technologies, dominance of the network by any set of stakeholders needs to be kept in check, lest openness gives rise to this dominance. Whereas closed modes of discovery may favour incumbents, openness may also give rise to new, or emergent configurations of power. In light of this, Proposition 2 is offered:

Proposition 2. The successful management of technology is fundamentally associated with the dispersion of power, whereby control over the research process itself (and its outcomes) is, and remains, accessible

At this current time, with R&D models, and particularly healthcare discovery models, at the mercy of the need for high levels investments under conditions of uncertainty about returns to these investments, the discovery process cannot be considered to be entirely accessible, and the outcomes of such a process are therefore also unequally distributed. Pharmaceutical investments, for example, will be skewed towards wealthy markets, and poorer populations will typically be disadvantaged if there is no mechanism through which firms can obtain returns on investment without targeting only markets wealthy enough to recoup investment costs. Private firms, however, can take advantage of openness to lower their costs of R&D, but may have few incentives to do so if market power is concentrated.

Arguably, at the extremes of low or high openness, and of low and high power relationships, the societal impacts of technology will be very different, and it might be in conditions of high openness and dispersed power relationships that collaborative networks of human stakeholders would have an improved ability to manage rapidly increasing technological change, and to more effectively mitigate its threats. To better ground the propositions derived here in relation to the scenarios they predict, and the scenarios associated with their opposite extremes, Fig. 2 relates the extremes associated with each proposition. The extreme states of Proposition 1, namely openness as a mode of discovery versus closed modes of discovery, are related to the extremes associated with Proposition 2, or the dispersion of power versus its opposite orientation, the intensification of power. Four modes of discovery are
taken to result. These four modes are now discussed.

4.1. Innovation closure

Conditions associated with closed modes of discovery and relatively high power dispersion are taken to be associated with a state of innovation closure, or a failure to make dramatic breakthroughs in important socially important areas. This is the state predicted by probabilistic innovation theory, whereby innovation failure, or gridlock persists on account of a failure to taking advantage of the exponentially increasing economies of scale in data analysis that are currently offered by technologies that already exist (Callaghan, 2016). Some have argued that in pharmaceutical innovation, for example, returns on investment have been stagnant for decades now. Although power is concentrated in markets, and innovation outcomes are inequitably distributed, the monopoly structure does not explicitly shut out new entrants, and the discovery system is considered to be open to disruption. This is broadly considered to reflect the current state of affairs. Because there is no explicit closure of the discovery process, the outcomes of discovery might be considered to be probabilistically related to investments in the discovery process. In others words, there is investment risk associated with innovation investments, but this risk can largely be quantified. Investment in innovation is not fundamentally uncertain in its outcomes.

4.2. Dystopian control

Under conditions of power intensification, the resources that dictate relationships within modes of discovery, and the outcomes of the discovery process, are within the power of certain agents, typically industry incumbents, or elites, because closed modes of discovery are expected to allow for the control of knowledge, and also its outcomes. Dystopian control is taken to represent a mode of discovery associated with high power differentials and low levels of openness. Under these conditions, inequality in discovery outcomes and in access to the discovery process is maximized. The power of knowledge creation is in the hands of elites, and both human progress and the threat of technological advancement are held in check, but at great cost to disadvantaged populations who are denied the benefits of innovation. This is arguably a feasible outcome if Joy’s (2000) strategy of technological relinquishment, or abandonment were adopted, as those less committed to it would not relinquish, and in so doing might increase their power over others.

4.3. Captured future

Under conditions of openness with high power relationships, it is still possible that industry incumbents, or new emergent groups might take control of the discovery process, in that openness might not be a sufficient condition for optimum effectiveness in the management of discovery. Given the efficiencies of shared knowledge, the consequences of concentrations of power in a context of openness, which facilitates the disruption of business models, are uncertain. Under conditions of such uncertainty, it may simply be not possible to calculate risk. Under the uncertainty associated with this quadrant, there might therefore be a shift in this quadrant toward any of the other three quadrants.

Opportunities for Internet-based global trade in goods and services and the exploitation of “unlimited, additional benefits” may result from AI inventions, but these “vast opportunities” for trade and productivity improvements need to be considered in relation to “dangers and disadvantages in terms of increased employment and greater wealth inequalities” (Makridakis, 2017:46). These advances may conceivably result in what Kurzweil (1999) has termed singularity, where nonbiological intelligence matches that of humans, and distinctions between human, machine, real reality and virtual reality disappear. Given the high levels of uncertainty associated with this mode of discovery, these outcomes need to be carefully considered.

Indeed, there might come a time where computers will choose those who serve in public office, given the poor choices humans often make in this area (Makridakis, 2017). As with the heralded advent of driverless vehicles, under conditions of openness and high power knowledge advantages that can be seized by the most powerful, technological change will be expected to accelerate, and attempts to manage it may be thwarted by powerful elites, perhaps in the form of a commercial arms race as technological advances fuel the pursuit of profitability. It is this mode that perhaps best captures the spirit of Joy’s (2000) criticism of material progress as a cause of the problem of dangerous technological advancement itself. Joy’s solution of relinquishment, however, might simply result in a shift toward dystopian control, as it is unlikely that elites will relinquish power. Under conditions of high openness coupled with dispersed power relationships, on the other hand, the mode of discovery might be uniquely suited to more effective management of societal problems, including that of dangerous technological change.

4.4. Age of effectiveness

The mode of discovery associated with a high level of openness and a high level of power dispersion is termed the ‘age of effectiveness’ as it is taken to offer the conditions most likely to contribute to the effective management of technologies. Digital technologies have “rendered new opportunities for learning that transcend barriers of time and space,” and harnessing the potential for robots as social agents in synergistic human-robot learning exchanges is distinct from many descriptions of AI which relegate humans to a “secondary role in the learning community” (Bricout et al., 2017:92). What such conceptions suggest is that technological advances can be harnessed in support of human learning and human agency in a world of AI. Advances in AI learning capabilities themselves show a dramatic increase over time. Milestones in this process include the reading of handwriting digits by
the neural network device (1990), vision-based navigation (1993), the development of speech (1998), and self-driving cars (2009) (Makridakis, 2017).

How then could human connectedness leverage human problem solving ability to the point at which it would be up to the challenge of effectively managing the complexities and dangers inherent in technological advancement and proliferation? Monat (2017) argues that the current level of human interconnectedness is growing, but in terms of ‘connections’ or ‘synapses’ is still short of the number of these connections in the human brain. He suggests that collective intelligence is emergent, in much the same way as the connections in an individual’s brain exhibit ‘emergent’ intelligence. He offers the notion that the brute number of connections in a human brain account for an individual’s intelligence, and that if the brute number of human connections in the world matched this number of brain connections then collective human behaviour would relatively be as intelligence as an average human. Although only a useful analogy, this notion suggests that collective intelligence might offer useful opportunities to leverage emergent human intelligence in the quest to manage problems like technological change.

If there are billions of people, however, why then has the world seemingly not developed more collective intelligence (currently about that of a chimpanzee, according to Monat, 2017)? He argues that this is because there are too few nodes (individuals that are connected), and there are too few connected to the Internet or news media globally, and because much information, if not biased or sensationalised, is filtered by the media. According to Monat (2017:27):

A fundamental difference between humans and other animals is that humans are highly self-aware while other complex animals are less so; and simple creatures like mosquitoes are not self-aware at all. Some researchers believe that self-awareness is an emergent property of a complex neural network. If this is so, then high self-awareness should appear when a neural network approaches the complexity of the human brain (∼90 billion neurons and 10^{14} synapses). If one takes a much broader view and considers all of humanity as a neural network, then today there are ∼7 billion individual elements, of whom ∼3 billion are interconnected via computers, smart phones, tables, and the Internet. By morphological analogy, as human interconnectivity continues to grow and strengthen, eventually humanity will approach ∼70 billion interconnected humans, at which point we will become highly self-aware as a single human super-organism. This organisational self-awareness may manifest as the elimination of wars, hunger, and strife, and as the collaboration of all individual elements working together for the greater good of humanity.

The lesson that emerges from this concept is that it is human collaborations and working together that might be key to leveraging human problem solving abilities, in the form of collective intelligence. According to Nielsen (2012), innovations in the discovery process can amplify human collective intelligence. The notion that humans can only stay ahead of the threats of technology by improving their ability to learn and manage technological change is associated not with technology pessimists, but with technology pragmatists.

The key argument of pragmatists is that by focusing on intelligence augmentation the dangers of AI can be managed, while “providing the means to stay ahead in the race against thinking machines and smart robots” (Makridakis, 2017:52). Some pragmatists have argued that AI technologies can be controlled using OpenAI together with regulation, as open systems that are not hidden behind proprietary doors will inherently offset risks (Peckham, 2016). High openness and high power dispersion might create the best conditions for humans to be able to manage technology, but this will necessitate taking advantage of technology itself to do this.

Humans may indeed have creativity advantages over intelligent machines. According to Jankel (2015:1), artificial intelligence has “raced forward in the last few years, championed by a libertarian, tech-loving and science-driven elite,” or “transhumanists who pronounce the eventual victory of the machine over nature.” He argues, however, that the belief that human brains are computers is “rooted more in metaphor than reality,” because algorithms act according to rules, and creative human disruptive innovators typically break rules, as breakthroughs are, by their nature, unpredictable; breakthrough “creativity is fundamentally organic, not algorithmic” (Jankel 2015:1). Within the next twenty years, however, rapid developments in AI are expected to result in breakthroughs based on deep learning that reflects the way children learn. Creativity might therefore not ultimately be the exclusive domain of humankind.

There is no limit to deep learning, on account of three factors, namely (i) open source software makes progress available to all and encourages the development of more powerful algorithms and cumulative learning, (ii) deep learning algorithms will use memory to apply problem solving to new contexts, and (iii) intelligence programmes will themselves write new programmes (Makridakis, 2017). According to Bricout et al. (2017:91) assisted technologies in the form of socially assistive robotics (SAR) can augment learning and action, and human-robot learning communities can develop, the success of which is contingent upon “how human users engage the networking capacity” of those communities. Thus, in the future, this level of machine intelligence might be unavoidable, and the key to successfully negotiating such an environment might be the ability we have to utilize technology to leverage human management capabilities.

Some might find these ideas unpalatable, given that they draw from a literature that engages with problems that are not yet part of our everyday experience. However, the fact that certain problems are wicked (Sardar, 2010) makes it necessary to confront them, as a consideration of future scenarios can help better manage these issues in the present. The arguments considered here are considered far future arguments. Baum (2015) argues that the far future argument, that “people should confront catastrophic threats to humanity in order to improve the far future trajectory of human civilization,” is important, notwithstanding the lack of motivation many have to do so, given their overriding concern for the near future rather than the far future, and the fact that there is little likelihood that they will experience the far future.

Can a technological future be a meaningful place for human life? Bricout et al. (2017:102) invoke Amartya Sen’s notion of capabilities relating to freedom, choice, and ability to act, to highlight the potential impact of vertical integration of technologies, or of a nexus future with universal accessibility in which the flow of information is unchecked. This future would give rise to “major
ethical concerns of users around confidentiality, privacy and autonomy,” and therefore human capabilities (p. 102). Again, these potentialities might be a function of the extent to which technological advances can be successfully managed. Many of the changes, however, may be difficult to negotiate. An example is the effect of AI and computerisation on the nature of human work, which also requires the effective management of technological change.

Using an analytic Markov chain model, Kim, Kim, and Lee (2017:6) analysed the effect of advances in big data, machine learning and robotics that have reduced human employment opportunities, concluding that “even if computerization proceeds at an uncontrollable pace and renders all previously non-susceptible jobs susceptible, a healthy portion of the future economy will consist of new jobs that permit a peaceful coexistence between humans and machines.” Kim et al. (2017) however caution that their results demonstrate that “legal and social limitations on computerization are key to ensuring an economically viable future for humanity.” Therefore controlling the crossover rate of occupations between susceptible and non-susceptible states “will help reduce the proportion of susceptible occupations in the economy (p.6).”

Kim et al. (2017:8) also suggest that with regard to employment loss due to technology, the “most viable solution for long-term success, however, may be a large-scale revision of the education system, in order to better equip future employees with the skills that will be necessary in a human-machine hybrid economy.” It is argued here that an age of effectiveness is perhaps possible, as long as openness is used to increase connectivity and collaboration between humans, which might allow collective intelligence to be used to leverage human management capabilities. It is also argued that human agency is also key to this challenge, and that there are ways to meet these challenges, but these might require action in the present. A careful consideration is necessary now, to understand how the education system, for example, and other human systems, can be reconfigured to meet these future challenges. Table 1 summarises concepts derived from the discussions above, and relates certain key challenges to each of the technological scenarios, or modes of discovery identified in Fig. 2. Further discussion of these relationships is offered in this table.

In summary, it is argued that at high levels of openness and high power dispersion, the low concentration of power over the discovery process is expected to enable effective management of discovery. This is considered a probabilistic era as outcomes can be calculated as risk. Dispersed power relationships mitigate the control and the power inequality threats. Destructive empowerment in the form of harmful activities are more effectively managed using the enhanced response capabilities associated with openness and distributed networks of collaborators. Accelerated problem solving may result under systems of collaborative problem solving, with lower power asymmetries providing the ability to harness the economies of scale of collective problem solving. Under conditions of technological change, which might be unforeseen, and therefore difficult to predict, it is arguably this quadrant which provides the most effective response to these potential dangers. Threats related to resource competition are also perhaps more effectively managed by the approach described by this quadrant. Arguably, the self-reproduction of artificial intelligence is a state that is subject to the extent to which technological change can be managed, and it is also this quadrant that provides the most useful approach to this. Similarly, the societal changes that influence human work are also, to some extent, a function of the effectiveness of the management of these changes.

5. Limitations

Certain limitations of this work need to be acknowledged. This article sought primarily to provoke further engagement with certain issues surrounding the development of dangerous technologies and their ultimate societal impact. The analysis undertaken here is however based on a critical review of literature, and therefore premised on subjective judgements of what aspects to prioritise in discussions. What insights were gained by the choices made here were necessarily at the cost of that lost by not considering other aspects. For example, anchoring the work on discussions of transparency and accountability as aspects of openness was based on the growing literature on responsible innovation, which was given priority. The choice to prioritise these perspectives was taken due to their accordance with the primary arguments of technology futures experts such as Bostrom and Tegmark. Given the need to subjectively provide an ordering, according to importance, of ideas and theory in this area, the analysis sought to draw on only what was taken to be the most salient work. In so doing, the analysis also provides insights that are at a certain level of abstraction.

To cover the necessary conceptual ground it was necessary to sacrifice depth of discussion in certain areas. Future work might address these deficiencies. Consideration of the six primary technological threats also necessitated subjective decisions as to which threats to recognise as primary and which to relegate to within-threat discussions of others. Given the uncertainties associated with attempts to make sense of technological futures, further work is invited, to improve on the categorisations made here. Indeed, it is hoped that further conceptual and data-driven work will draw out more detailed causal relationships between these threats (and highlight others), and ultimately provide tests of the predictions of the theoretical framework.

6. Conclusions

Given the substantial promise of technological advancement for the improvement of human lives, and given the threat of the proliferation of dangerous technologies, the objective of this paper was to offer certain insights for how these threats could be better managed. Certain key threats associated with the future proliferation of technology were identified. A theoretical model was developed, on the basis of theoretical propositions derived from the literature. Using these propositions as a heuristic frame, four future scenarios were identified, predicting different societal outcomes for different permutations of openness and the power of elites. Under conditions of high power and low openness, it was predicted that powerful elites might control innovation at the expense of relatively less powerful populations. The current global state of discovery was considered to be categorised by a mode of low power and low openness, associated with innovation gridlock, whereby few have access to the discovery process and slow innovation, particularly
| Key challenges                        | Technological Scenarios                                                                 |
|--------------------------------------|-----------------------------------------------------------------------------------------|
|                                      | Dystopian Control (probabilistic era)                                                   |
| Ethical issues in research            | Concentration of power through ‘behind closed door’ modes of discovery. Lack of ethical oversight and no way to enforce it. |
| Incentives in the research ‘industry’ | Private incentives dominate, with few incentives for societally important discovery.     |
| Formulaic research                   | Maximized formulaic structure of academic research, as traditions and silo approaches crowd out novel methodologies and innovations in the research process itself. |
| Tensions between gatekeeping and breakthrough research | Tacit fluency becomes the currency of discovery, as uni-disciplinary vocabularies enable closed systems of discovery. Strong normative gatekeeping in science, stifling innovation that does not favour incumbent powerful elites. |
| Status quo versus change             | Strong and active support for status quo by powerful incumbents- shutting down disruptive innovation and maintenance of monopoly conditions in discovery. |
| Status quo versus change             | Strong and active support for status quo by powerful incumbents- shutting down disruptive innovation and maintenance of monopoly conditions in discovery. |
| Distribution of outcomes from the discovery process | Inequality in outcomes | Potential inequality in outcomes, but uncertain | Inequality in outcomes, but uncertain |

**Table 1**
Technology scenario futures: key challenges and implications for society.

| Key challenges                        | Technological Scenarios                                                                 |
|--------------------------------------|-----------------------------------------------------------------------------------------|
|                                      | Captured Future (uncertainty era)                                                      |
| Ethical issues in research            | Openness allows ethical oversight that might mitigate a greater potential for technological change on account of openness. |
| Incentives in the research ‘industry’ | Incentives are emergent, and uncertain, as rapid technological change from openness can be captured by more powerful groups. |
| Formulaic research                   | Powerful incumbents or emergent elites resist novel ideas and seek to use technological knowledge provided by openness to establish monopoly or oligopolistic conditions. |
| Tensions between gatekeeping and breakthrough research | Innovation spurred by openness is not independent of power contests and dynamics can shift toward the characteristics of any three of the other quadrants. |
| Status quo versus change             | Openness mitigates against the status quo but powerful lobbies can steer contextual change in the direction of either of the other three scenarios. Outcomes are uncertain. |
| Status quo versus change             | Openness mitigates against the status quo but powerful lobbies can steer contextual change in the direction of either of the other three scenarios. Outcomes are uncertain. |
| Distribution of outcomes from the discovery process | Inequality in outcomes | Potential inequality in outcomes, but uncertain | Inequality in outcomes, but uncertain |

| Key challenges                        | Technological Scenarios                                                                 |
|--------------------------------------|-----------------------------------------------------------------------------------------|
|                                      | Innovation closure (probabilistic era)                                                  |
| Ethical issues in research            | Current paradigm of research and R&D in which a lack of innovation and progress against societal outcomes is unethical in itself. |
| Incentives in the research ‘industry’ | Private incentives predominate, and societally important innovations do not keep pace with private business innovations (current status quo). |
| Formulaic research                   | Current divergence of academic fields, some deepening a silo focus, and others seeking support and resources from close linkages with practitioner communities. Self-preservation logics dominate, as in current system of discovery. |
| Tensions between gatekeeping and breakthrough research | Field-specific value added. A focus on practitioner field requirements |
| Status quo versus change             | Anti-formulaic ethos, whereby novel methods are embraced, while institutional structures provide sustainability to research, reconceptualised as discovery endeavours. |
| Status quo versus change             | Anti-formulaic ethos, whereby novel methods are embraced, while institutional structures provide sustainability to research, reconceptualised as discovery endeavours. |
| Distribution of outcomes from the discovery process | Inequality in outcomes | Potential inequality in outcomes, but uncertain | Inequality in outcomes, but uncertain |

| Key challenges                        | Technological Scenarios                                                                 |
|--------------------------------------|-----------------------------------------------------------------------------------------|
|                                      | Age of effectiveness (probabilistic era)                                                |
| Ethical issues in research            | Maximized transparency and accountability together with low information and power asymmetry reflect a strong ethical engagement. |
| Incentives in the research ‘industry’ | Incentives are balanced, with institutions developing to maintain balance between private and societally important innovations due to more effective global management of discovery itself. |
| Formulaic research                   | Anti-formulaic ethos, whereby novel methods are embraced, while institutional structures provide sustainability to research, reconceptualised as discovery endeavours. |
| Tensions between gatekeeping and breakthrough research | Problem-centric focus- innovativeness and societal contribution as gatekeeping criterion |
| Status quo versus change             | Disruptive innovation and transparency resulting in outputs becoming real time inputs into discovery process. Innovative management of problems and threats, with speedy disaster response. |
| Status quo versus change             | Disruptive innovation and transparency resulting in outputs becoming real time inputs into discovery process. Technological advances are matched by innovative management of problems and threats, with speedy disaster response. |
| Distribution of outcomes from the discovery process | Potential for equitable outcomes from discovery | Potential for equitable outcomes from discovery | Potential for equitable outcomes from discovery |
healthcare innovation, results in high inequality in outcomes, as only wealthier markets attract substantial R&D investments from firms. Under conditions of high power and high openness, however, the consequences of technological advancement and proliferation were taken to be uncertain, as the discovery process might be dominated by powerful elites who have the power to either curtail innovation or enable the proliferation of dangerous technologies. It was finally argued that conditions of high openness and high power dispersion might be optimal for the development of the management capabilities required to successfully manage technological change, and that technology itself may hold the key to developing these capabilities. According to this pragmatic perspective, an important avenue for future research is how collective intelligence might be leveraged using technology, as this might offer a useful approach to keeping pace with machine intelligence and other threats associated with a technological future. Ironically, it is typically only in the face of a common threat that humans become united, and seek radically improved collaborations. Uniting now, in the present, to develop radically enhanced collaborative capabilities might be our saving grace, and it is the responsibility of future studies to lead the way.

Although recommendations were previously made throughout the above discussions, certain overarching recommendations derive from the analysis. First, given that the trial-and-error paradigm of practice and management may produce excessive risk, a new formalised approach may be necessary, and this is recommended. If a single mistake can have catastrophic consequences, then technology safety research should be an integral part of the technology development process. This may require openness and an explicit focus on the dispersion of societal power relationships so as to empower responsible innovation that is transparent in terms of its dangers.

Second, an ordering of the threats considered here suggests that a re-ordering of the roles of societal stakeholders is necessary. In addition to increasing engagement and the application of principles of responsible innovation, technology safety researchers as a stakeholder group need to be empowered and integrated into debates and practice related to dangerous technologies. Structural changes in academia are recommended, to increase formalised support and funding for technology futures and technology safety research.

Third, given recent literature that suggests that technological innovations applied to the research process itself can enable economies of scale without compromising rigor, the incorporation of these innovations into technology safety research is recommended. It is necessary to become more proactive rather than reactive, investing in safety research immediately. This is perhaps a new way of thinking about how we do research on, and how we develop, dangerous technologies. In line with this new way of thinking, research into technological futures should be intensified. This research should provoke engagement and prompt more ‘voices to be heard.’ The futures literature is well placed to lead the way with this.

In terms of surviving a technological future, Bostrom’s (2017:300) words perhaps bear repeating here, in that the “most appropriate attitude may be a bitter determination to be as competent as we can, much as if we were preparing for a difficult exam that will either realize our dreams or obliterate them.” Indeed, as Sardar (2010) puts it, futures studies are futureless, in that their contribution is to the present. By interrogating wicked problems related to technological proliferation, and formulating appropriate alternative technological scenarios, we will be better able to prepare for them in the present, otherwise certain of these less desirable scenarios may ultimately come to characterise the present.

References

Baum, S. D. (2015). The far future argument for confronting catastrophic threats to humanity: Practical significance and alternative. Futures, 72, 86–96.

Bernstein, A., Klein, M., & Malone, T. W. (2012). Programming the global brain. Communications of the ACM, 55(5), 41–43.

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., et al. (2009). Public participation in scientific research: Defining the field and assessing its potential for information science education. A CAISE inquiry group report. Washington: Centre for Advancement of Informal Science Education.

Bostrom, N. (2017). Superintelligence. Paths, dangers, strategies. Oxford: Oxford University Press.

Bricout, J. C., Sharma, B. B., Baker, P. M. A., Behal, A., & Boloni, L. (2017). Learning futures with mixed sentence. Futures, 87, 91–105.

Brynjolfsson, E., & McAfee, A. (2011). Race against the machine. Lexington: Digital Frontier Press.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York: W.W. Norton & Company.

Callaghan, C. W. (2016). Disaster management, crowdsourced R&D and probabilistic innovation theory: Toward real time disaster response capability. International Journal of Disaster Risk Reduction, 17, 238–250.

Carbonell, J., Sánchez-Esguerra, A., & Carro, B. (2016). The role of metaphors in the development of technologies. The case of artificial intelligence. Futures, 84, 145–153.

Cohen, J., & Malakoff, D. (2012). On second thought, flu papers get go-ahead. Science, 336(6077), 19–20.

Doughlas, H. (2000). Inductive risk and values in science. Philosophy of Science, 67(4), 559–579.

Foucault, M. (1982). The subject and power. New York: W.W. Norton & Company.

Freeman, R. E. (1984). Strategic management: A stakeholder approach. Boston: Pitman.

Funtowicz, S. O., & Ravetz, J. R. (1994). Uncertainty, complexity and post-normal science. Environmental Toxicology and Chemistry, 13(12), 1881–1885.

Gilding, M. (2002). Families of the New Millenium. Family Matters, 62, 4–10.

Grunwald, A. (2011). Responsible innovation: Bringing together technology assessment, applied ethics, and STS research.

Kupferschmidt, K. (2017).

Kim, Y. J., Kim, K., & Lee, S. (2017). The rise of technological unemployment and its implications on the future macroeconomic landscape. Futures, 87, 1–9.

Kurzweil, R. (1999). The age of spiritual machines. New York: Penguin.

Kurzweil, R. (1999). The age of spiritual machines. New York: Penguin.
