Journal of Futures Studies
2021,Vol. 26(2) 51–65
DOI: 10.6531/JFS.202112_26(2).0004

Article

Brain Computer Interfaces: A New Existential Risk Factor

Jack Rafferty

Co-Founder & Director at LEEP, London, England, United Kingdom

Abstract

This paper identifies a new existential risk factor that has not been recognised in prior literature: Brain-Computer Interfaces (BCIs). We illustrate how BCI technology could significantly raise the existential risk from global totalitarianism in the near future. In particular this is achieved not just by expansion of surveillance, but the expansion of brain stimulation. At present, this risk factor has been entirely unnoticed. We suggest that given the high likelihood of its impact, and the possible magnitude of such an impact, it deserves more attention, more research, and more discussion.

Keywords

Existential Risk, Brain-Computer Interface, Totalitarianism

Introduction

Brain-Computer Interfaces (BCIs) are technologies that allow the brain to interface directly with an external device. In particular, BCIs have been developed to read mental and emotional content from neural activity and have been developed to intentionally stimulate or inhibit certain kinds of brain activity. At present BCIs are used primarily for therapeutic purposes, but their potential use case is much wider.

Though this sounds somewhat like science fiction, the current state of the technology is much more advanced than most realise. In particular, well-corroborated research prototypes already exist (Guenther et al, 2009; Moses et al, 2019); a number of companies, including Facebook and Neuralink, are working to commercialise this technology over the coming decades (Constine, 2017; Musk, 2019); and there is widespread agreement among BCI researchers (86% agreement) that this technology is not just theoretically feasible, but will be on market in the near future (Evers & Sigman, 2013; Merrill & Chuang, 2018; Nijboer et al, 2011). The risks this technology poses however, have been almost entirely neglected.

This paper outlines how the development and widespread deployment of BCIs could significantly raise the likelihood of long term global totalitarianism. We suggest two main methods of impact. Firstly, BCIs allow for an unparalleled expansion of surveillance, as they enable states (or other actors) to surveil even the mental contents of their subjects. Secondly, BCIs make it easier than ever for totalitarian dictatorships to police dissent by using brain stimulation to punish dissenting thoughts, or even make certain kinds of dissenting thought a physical impossibility.

**Corresponding author*.
E-mail addresses: jack2rafferty@gmail.com (Jack Rafferty)

Received: 12 August 2020 and Accepted: 05 January 2021

Definitions

Existential risk

Global existential risks are risks that threaten the premature extinction of earth originating life (Bostrom, 2012) or that threaten "the permanent and drastic reduction of its potential for desirable future development" (Cotton-Barratt & Ord, 2015). As such, while some existential risks pose the danger of extinction, some do not. Nuclear war is an example of an existential risk that poses a risk of human extinction. However, irreversible global totalitarianism is often considered an existential risk too, because even without posing any extinction risk, it has the capacity to irreversibly destroy or permanently lessen a great level of humanity's potential (Ord, 2020).

Risk factors & security factors

A risk factor is typically not an existential risk, but is something that makes an existential risk more likely. Often this means increasing the probability that a certain event might occur or increasing the likelihood that a catastrophic event becomes existential. For example, relating to risk of extinction from nuclear war, a risk factor might increase the likelihood that a nuclear war begins. Alternatively, a risk factor might instead increase the likelihood that a nuclear war is catastrophic enough that it leads to extinction or unrecoverable collapse. Both would be valid risk factors.

A practical example of a risk factor is international conflict. While international conflict could have very negative impacts, by itself it would not be considered an existential risk, because conventional war offers little chance of extinction or of permanently curtailing our long-term potential. However, higher levels of conflict between great powers might drastically raise the odds of a nuclear war and thus increase the odds of an existential catastrophe. As such, high levels of international conflict could be considered a risk factor for nuclear war.

While risks and risk factors are often separate categories, this is not universal – the categories are not mutually exclusive. For example, Torres (2016) argues that climate change could be both an existential risk in itself, as runaway global warming could make the earth physically uninhabitable and threaten extinction, but could also be a risk factor that might increase the likelihood of other kinds of existential risks. For example, increased climate change might increase global instability, make international cooperation less viable, and compromise the conditions necessary to address other existential risks. Designing, building, and launching solutions to risks such as an asteroid strike would be far more difficult in a world preoccupied with other survival issues.

A security factor is the opposite of a risk factor. It is something that reduces the chance that an existential risk occurs. For example, good international governance may be a security factor that lessens the chance of nuclear war.

Just as action to avoid existential risk is crucial, dealing with risk factors can be as important or in some cases even more important than dealing with the risks themselves (Ord, 2020). For example, if the chance of a particular existential risk occurring is 10%, but a risk factor brings this chance up to 90%, it may end up being more cost effective to address the risk factor before addressing the risk itself. This is not always the case, but there can be strong justification for working on alleviating existential risk factors when it is cost effective to do so.

This paper seeks to identify and outline the danger and likelihood of a new and unnoticed existential risk factor.

Brain-Computer Interfaces

Outline of current brain-computer interfaces

A brain-computer interface (or BCI), is an interface between a brain and an external device. Certain forms of BCIs already exist; the term refers to a range of technologies, used for a number of purposes. At present, the most well known commercial uses of BCIs include recovering lost senses, as with cochlear implants used to restore hearing and retinal implants to restore sight (Anupama et al, 2012). However, BCIs have a vastly broader set of uses that already exist as either in-use medical technologies or as well corroborated research prototypes. This section outlines a few of these uses to give an idea of the current and near term scope of the technology.

For the purposes of our explanation, there are two broad functions of BCIs. The first kind of BCIs are able to read neural activity, record it, interpret it, send it, and use it for a variety of purposes. The second kind of BCIs are able to write to the brain. They are able to influence and modify brain activity, stimulating or suppressing various responses by using skull mounted micro-electrodes, or by using less invasive transcranial electrical stimulation. These two types could be combined and used together, though for clarity we will refer to them as type 1 and type 2 BCIs to differentiate function.

Type 1 BCIs are able to read neural data, but also report and send this data for a number of purposes. These have already been used to translate speech from neural patterns in real time (Allison et al, 2007; Guenther et al, 2009; Moses et al, 2019), and to detect positive and negative emotional states from neural patterns (Wu et al, 2017). It is expected that near term BCIs of this kind will be able to detect intentional deception, detect even subconscious recognition, and detect more precise and complex thought content (Bellman et al, 2018; Bunce et al, 2005; Evers and Sigman, 2013; Roelfsema, Denys & Klink, 2018). There are many practical uses of recording and interpreting neural data. So far, BCIs have been used in primates to allow them to control prosthetic limbs and smart devices with thought, by sending mental commands directly to the relevant device (Carmena et al, 2003; Ifft, 2013; Moore, 2003). These same techniques have also been used to assist people who are paraplegic or quadriplegic by providing them with a neural shunt that records messages from the brain and sends these messages directly down to where the muscles are activated, allowing patients to use previously disabled limbs (Moore, 2003). Many companies also have the long term goal of allowing users to mentally transmit messages to other BCI users, allowing silent communication with only a thought (Kotchetkov et al, 2010).

The uses of type 2 BCIs are even more varied. Many uses are therapeutic. Deep brain stimulation for example, has used neural stimulation to treat various disabilities and conditions, including Parkinson's disease (Deuschl et al, 2005; Glannon, 2009; Perlmutter, 2006). Similar techniques have been used to alleviate disorders such as OCD (Abelson et al., 2005; Greenberg, 2006), and have been suggested as potential future treatments for conditions like Alzheimer's and depression (Laxton et al., 2013; Mayberg et al., 2005), and even to restore function in those with motor disability after a stroke (Gulati et al, 2015).

Through deep brain stimulation, control of physical pain responses is also a possibility. Such techniques have been used to alleviate chronic pain (Bittar et al, 2005a; Kumar et al, 1997), treat phantom limb syndrome (Bittar et al, 2005b), augment memory (Suthana, 2012; Hamani et al., 2008), and more. Just as BCIs can currently suppress pain, pain responses can also be stimulated for a variety of purposes, from interrogation to incentivisation to punishment. Similarly, BCIs are already able to artificially stimulate or suppress emotional reactions (Delgado, 1969; Roelfsema et al., 2018). These are just a few of the corroborated functions of BCIs. In future, it has been suggested that BCIs could be used as a possible treatment for cravings and addictions, and as a way to alter

internal drives and rewards systems (Mazzoleni & Previdi, 2015; Halpern, 2008).

"Consider eating a chocolate cake. While eating, we feed data to our cognitive apparatus. These data provide the enjoyment of the cake. The enjoyment isn't in the cake per se, but in our neural experience of it. Decoupling our sensory desire from the underlying survival purpose [nutrition] will soon be within our reach." – Moran Cerf, Professor at Northwestern University, Employee at Neuralink.

Future brain-computer interfaces

The potential uses of BCIs are well corroborated. The primary difficulties at present include scaling down costs, size, and invasiveness, and scaling up precision to allow BCIs to target more neurons, more specifically.

At present, significant research and development is being done on BCIs to address these issues, to expand their capabilities, and to make BCIs orders of magnitude cheaper, more precise, less invasive, and more accessible to the broader population. Companies currently working on developing cheap, publicly accessible advanced BCIs include Facebook (Constine, 2017), Kernel (Kernel, 2020; Statt, 2017), Paradromics and Cortera (Regalado, 2017), and Neuralink (Musk, 2019). In addition to this, DARPA, the research arm of the US military, is funding significant research in this direction (DARPA, 2019; Kotchetkov et al, 2010), as is the Chinese government (Munyon, 2018; Tucker, 2018). In short, with so many well funded companies and governments working on this problem, it is likely that these barriers will quickly fall.

To reinforce this, market trends for BCIs speak of strong expected growth. The global BCI market was valued at $1.36 billion in 2019 but is projected to reach $3.85 billion by 2027, growing by 283% in just eight years (Gaul, 2020). The likelihood of development of X-risk relevant BCIs within this century is relatively high.

Not all BCIs involve 'humanity' scale risk

As a point of clarification, this paper does not argue that all BCIs act as an existential risk factor. It seems incredibly unlikely that cochlear implants have any impact on the likelihood of any existential risk. However, we do argue that certain kinds of more advanced BCI may be extremely dangerous and may drastically raise the risk of long-lasting global totalitarianism.

Current Literature on Risks from BCIs

Previously identified risks

The current literature on global existential risk from BCIs is scarce. The vast majority of the literature on risk from BCI has focused on impacts at a very low scale. Such low-scale risks that have been considered include surgical risk from operations, possible health related side effects such as altered sleep quality, risk of accidental personality changes, and the possibility of downstream mental health impacts or other unknown effects from BCI use (Burwell et al., 2017). Potential threats to individual privacy have also been identified – specifically, the risk of BCIs extracting information directly from the brains of users (Klein et al, 2015).

At a higher scale, Caplan (2008) successfully identified 'brain scanning technology' as a factor that may impact existential risk at some point in the next thousand years by assisting with the maintenance of dictatorships. However, Caplan focuses only on risk within the next millennium,

and does not consider the high potential for this to occur in a far shorter time frame; in particular, within the next hundred years. He also only briefly mentions brain scanning as a technology and does not consider the risk from brain scanning technology being present and active in all citizens at all times. Such widespread use is a stated goal of multiple current BCI companies. Finally, Caplan does not consider the full depth of the impact of BCIs – only mentioning the capacity of brain scanning to increase the depth of surveillance, while ignoring the existential risk posed by the widespread use of brain stimulation.

Cybersecurity and coercion

A final risk identified in prior literature is cybersecurity, though prior literature has primarily focused on the threat to individuals. Specifically, the risk has been discussed in relation to vulnerabilities in information security, financial security, physical safety, and physical control (Bernal et al, 2019a). BCIs, just like computers, are vulnerable to manipulation by malicious agents. BCIs and brain scanning offer an unprecedented level of personal information, passwords, as well as data about a user's thoughts, experience, memories and attitudes, and thus offer an attractive terrain for attackers. It is likely that security flaws will be used by malicious actors to assist with cybercrime. Further previously identified risks here include risk of identity theft, password hacking, blackmail, and even compromising the physical integrity of targets who rely on BCIs as a medical device (Ienca, 2016; Bernal et al, 2019b). The use of deep brain stimulation for coercion or control of BCI users is also a possible source of risk (Demetriades et al, 2010). Corroborated possibilities here include control of movement, evoking emotions, evoking pain or distress, evoking desires, and impacting memories and thinking processes – and these are just the earliest discovered capabilities (Delgado, 1969). However, past papers have exclusively focused on risk to individuals; that individuals may be sabotaged, surveilled, robbed, harmed, or controlled. Past research has not yet explored the risk to humanity as a whole.

This paper seeks to take the first steps to fill that gap and outlines the risks that BCIs provide at a broader, global scale, addressing the risk they pose to the future of all of humanity.

Higher Scale Risks: BCI as a Risk Factor for Totalitarianism

Risk from neural scanning: ability to surveil subjects

Dissent from within is one of the major vulnerabilities of totalitarian dictatorships. BCIs offer dictators a powerful tool to counteract this weakness. Increases in abilities for surveillance would make it easier to identify and root out dissent or root out skeptics who might betray the party, and thus would make it easier to maintain totalitarian control. While conventional surveillance may allow for a high level of monitoring, such as tracking of citizens' behaviour and actions, it provides no way for a dictator to peer inside the minds of their subjects. Because of this, the difficulty of identifying the attitudes of careful defectors remains high. BCIs constitute an unprecedented threat here. Surveillance through already existing methods may fail to expose some threats to a totalitarian regime, such as party members who carefully hide their skepticism. But BCI based surveillance would have no such flaw.

The level of intrusion here is potentially quite severe. With the advancement of BCIs, it is highly likely that in the near future we will see a rapid expansion in the ability to observe the contents of another's mind. Some researchers claim that advanced BCIs will have access to more information about the intentions, attitudes, and desires of a subject than those very subjects do themselves, suggesting that even subconscious attitudes and recognition, as well as intentional deception and

hidden intentions will be detectable by BCIs (Bunce et al, 2005; Evers and Sigman, 2013). Already, BCIs are able to detect unconscious recognition of objects that a subject has seen but cannot consciously remember seeing (Bellman et al, 2018).

Others have even suggested that by more precisely recording the activity of a larger number of neurons, future BCIs will be able to reveal not just perceptions and words, but emotions, thoughts, attitudes, intentions, and abstract ideas like recognition of people or concepts (Roelfsema et al., 2018). Attitudes towards ideas, people, or organisations could be discovered by correlating emotions to their associated thought content, and dictatorships could use this to discover attitudes towards the state, political figures, or even ideas. This would allow detection of dissent without fail and allow a dictator to quell rebellion before a rebellious thought is even shared.

Some might hope for BCIs that do not have this level of access, but accessing and recording mental states is a fundamental and unavoidable feature of many BCIs. In order to achieve their desired functions, many BCIs need a clear way to read neural data. Without significant neural data they simply cannot function – it is impossible to translate neural data to exert some function if one does not have access to that neural data. Brain stimulators and BCIs are specifically designed to allow this kind of access; it is crucial for the effective functioning of the device (Ienca, 2015). It is of course possible that BCIs made by some companies will be exclusively targeted to certain sections of the brain, for example, only targeting areas associated with speech, and not targeting other areas associated with emotions or thought. This is conceivable, though it is not clear that all companies and countries would do the same. Furthermore, the utility gained by expanding to other areas of the brain beyond the speech centre means it is highly doubtful the technology will remain restrained indefinitely.

It is likely that BCIs will be created by companies, which have strong financial incentive to record the neural states of users, if only to gain more information with which to improve their own technology. This information could be requisitioned by governments, as is frequently done to tech companies at present – even in democratic countries. Further exacerbating this problem, privacy laws have a history of struggling to keep pace with technological advancements. In more authoritarian countries, neural data might be transmitted directly to state records, and the preservation of privacy may not be attempted at all.

In essence, BCIs allow an easy and accurate way to detect thoughtcrime. For the first time, it will be possible for states to surveil the minds of its citizens. Deep surveillance of this kind would increase the likelihood that totalitarian dictatorships would last indefinitely.

Risks from brain stimulation: ability to control subjects

In addition to recording neural activity, there is an even greater threat that has not been considered as an existential risk factor in any prior literature. In addition to reading brain activity, BCIs are able to intentionally influence the brain. In particular, future BCIs will be able to rewire pleasure and pain responses and allow us to intentionally stimulate or inhibit emotional responses en masse. Where this is done consensually and is desired, this may be of some benefit. However, nothing about this technology guarantees consent.

In addition to being able to identify dissident elements more effectively than ever (due to increased surveillance), BCIs will also powerfully increase the ability of states to *control* their subjects, and their ability to maintain that control indefinitely. In such a situation, identification of dissidents would no longer be necessary, as a state could guarantee that dissident thought would be a physical impossibility. Finely honed BCIs can already trigger, and associate, certain emotions or stimuli with certain concepts (Roelfsema et al., 2018). This could be used to mandate desirable emotions towards some ideas or make undesirable emotions literally impossible. Though this

possibility has been discussed in literature for its therapeutic uses, such as triggering stimulation in order to respond to negative obsessive thoughts (nullifying negative emotions caused by such thoughts) there is huge potential for misuse. A malicious controller could stimulate loyalty or affection in response to some ideas, or even for specific organisations and people; and could stimulate hatred in response to others. It could also inhibit certain emotions, so that citizens would be physically unable to feel anger at the state. The ability to trigger and suppress emotional content with BCIs has already existed for years (Delgado, 1969). Combined with complex and detailed reading of thought content, this is a highly dangerous tool.

Some might argue that dissident action may be possible even with an outside agent controlling one's emotional affect. This is highly debatable, but even without any control of this emotional content, the risk from BCIs is still extreme. BCIs could condition subjects to reinforce certain behaviour (Tsai et al, 2009), or could be used to stimulate aversion to inhibit undesired behaviour (Lammel et al, 2012), or stimulate the pain or fear response (Delgado, 1969) and cause intense and unending pain in response to certain thoughts or actions – or even in response to a lack of cooperation. Even without controlling emotional affect, the state could punish dissident thoughts in real time, and make considering resistance a practical impossibility. This is a powerful advantage for totalitarian states, and a strong reason for authoritarian states to become more totalitarian. In addition to surveillance, it creates a way to police the population and gain full cooperation from citizens in a way that (once established in all citizens) could not be resisted. Machine learning programs scanning state databases of neural activity could detect thought patterns towards the state that are deemed negative and punish them in real time. Or, if the state is more efficient, it could simply stimulate the brains of subjects to enforce habits, increase loyalty, decrease a subject's anger, or increase their passivity (Lammel, 2012; Tsai et al, 2009). At worst, the brain could be re-incentivised, with errant emotions turned off at the source, so that dissenting attitudes are unable to ever form. Even high level dissent or threat of coup would be virtually impossible in a totalitarian state of this kind. Its long term internal security would be assured.

BCIs also offer an easy way to interrogate dissidents and guarantee their cooperation in helping to find other dissident camps – which might be otherwise impossible. In past resistances, certain dissidents have been considered near-impossible to completely wipe out due to features of terrain making it impossible to locate them in a cost effective way. If the government were able to access and forcibly apply BCIs, resistance would be a dramatically weaker obstacle. Dissenters might normally lie or not cooperate, but with BCIs, they simply need to be implanted and rewired. Then they would be as loyal and cooperative as any other, and could actively lead the state to their previous allies. Even unconstrained defectors could not be fully trusted as they may one day be controlled by the state.

Another issue for the long term survival of totalitarian dictatorships is coups or overthrows from within, as citizens or party officials are often tempted by different conditions in other states. With BCIs, the loyalty of regular citizens and even party officials could be assured. In current dictatorships, wiping out dissidents (particularly nonviolent dissidents) often has a significant social cost that can delegitimise and destabilise regimes (Sharp, 1973). A dictatorship whose citizens are all implanted with BCIs would not pay this social cost. At present, when dictators crack down it can cause riots and resistance, which can cause dictatorships to fall. With BCIs, governments will not need to appease their citizens at all to maintain loyalty. They need only turn up the dial.

Global Strategic Implications of BCIs

In this section we explore some global strategic implications of BCIs. In particular, that BCIs allow totalitarian regimes to be stable over the long term, even without requiring global totalitarianism.

We also argue that BCIs make authoritarian regimes more likely to become totalitarian in the first place, and that BCIs create a strategic equilibrium that inclines us towards a world where all countries become totalitarian.

Totalitarian states may fail for a few reasons. Conquest by external enemies is a danger, and since totalitarian states tend to stagnate more than more innovative liberal states, this may be a danger that grows over time. Internal dangers occur too; citizens may choose to rebel after comparing their lives to more prosperous countries in the outside world. Violent and nonviolent resistances have been able to overthrow even harsh authoritarian regimes (Chenoweth, 2011), and at least one totalitarian state has been overthrown by popular uprising (specifically, the Socialist Republic of Romania).

It has been suggested that the presence of successful liberal countries may tempt defection among the members of authoritarian and totalitarian countries. Maintaining the morale of citizens and the inner elite is a primary issue. Orwell (1945) and Caplan (2008) both propose that global totalitarianism would allow a totalitarian state to escape these risks of rebellion, as there would be no better condition for subjects to be tempted by or to compare their lives to. However, with BCIs, global totalitarianism would no longer be necessary; BCIs can disarm these issues. Not only is identification of dissent easier; the capacity for dissent can be entirely removed such that it never even begins. Loyalty and high morale can be guaranteed and biochemically enforced. Typically, it is hard to maintain commitment to totalitarian ideologies when free societies deliver higher levels of wealth and happiness with lower levels of brutality and oppression. BCIs could neutralise this problem, making temptation physically impossible, loyalty guaranteed, and regimes internally stable forever.

In addition to this internal stability, regimes could also be stable to external threats through the development of nuclear weapons, which powerfully discourage war and provide security from foreign nations. Being safe from both internal and external threats would have significant impacts on the lifespan of a totalitarian country.

In addition to increasing the longevity of dictatorships, BCIs also increase the likelihood that totalitarian systems will form. In particular, this is because conventional dictatorships will now have a far more powerful incentive to become totalitarian, as BCIs would make it cheap, easy, and most importantly, incredibly advantageous to do so. As established above, there is significant survival value in being able to identify and remove all opportunities for internal dissent and make rebellion literally ‘unthinkable’. If the misuse of BCIs would vastly improve the odds of survival for both a dictator and their government, then there is powerful reason for dictators to transition to BCI reinforced totalitarianism. Survival would be a powerful reason to descend to totalitarianism. It will be cheaper and easier to surveil and to police citizens than ever before, and the benefits to dictators of doing so will be large. Therefore, BCIs may increase not just the longevity of totalitarian states, but also the likelihood that they occur in the first place.

Finally, and most importantly, BCIs create a worrying strategic environment that may incline all countries to eventually become totalitarian, and may incline totalitarianism to entrench itself globally. With BCIs to stabilise themselves from internal resistance and nuclear weapons to stave off invasion, totalitarian countries would almost never fall. They would be secure from internal threats, and secure from external ones. Meanwhile, democratic countries that do not brainwash their citizens could be secure externally but might still at some point degenerate to a more authoritarian form of government. Democratic governments have rarely lasted more than a few centuries in history, and have often temporarily slid into dictatorship or authoritarianism.

At present, democracies can descend to dictatorship, and dictatorships can have revolutions and rise to democracy. With BCIs however, democracies can still collapse, but dictatorships are able to last forever. This is a dangerous equilibrium, as it means that free countries will still eventually fall,

as they do at present, but when they do, they will not be able to climb back out. Democracies could still collapse to dictatorship, but dictatorships could never rise from that state. In a world where democracies are mortal but dictatorships live forever, the global system is inevitably inclined towards totalitarianism.

Over time, one by one, individual democratic powers would fall to authoritarianism and use BCIs to establish irreversible, immortal totalitarian dictatorships in their own regions. The former-democracy would then be able to maintain a stable, BCI-reinforced authoritarian state indefinitely. In a world where a) countries that preserve mental freedom might possibly degenerate into such totalitarian countries, b) totalitarian dictatorships are immortal, and c) there is no available territory for new free countries to be founded, it creates an equilibrium whereby countries will steadily converge upon becoming dictatorships. A free country might not be free forever, and might at some point collapse into dictatorship, and then reinforce itself with BCIs. However, once a BCI reinforced dictatorship is established, it is likely to last indefinitely. Eventually, every free country would fall. And fallen countries would remain fallen forever. This provides a clear path into a multi-polar global totalitarian order.

At present, this is not a major concern, as dictatorships (even totalitarian dictatorships) can be overthrown. Thus the likelihood of all countries falling to totalitarianism one by one, without having countries rise back to democracy, is very low. We suggest that BCIs make the path to a global multi-polar totalitarian system much more likely.

The timeframe for such a shift is hard to predict. However, if BCI reinforced totalitarianism is already entrenched in a greater number of countries, then the problem may be drastically harder to stop, and the overall risk will be higher. This offers an unusual circumstance in regard to existential risks. With more time, it is likely that more countries will fall, and the more totalitarian countries there are, the harder this problem will be to solve. As such, this is a problem that may be easier to address earlier than later.

The possibility of anti-proliferation

There may be less of a risk from BCI use if there is a strong ability to prevent technological proliferation to despotic governments. This is conceivable. However, the level of success over the last 100 years at preventing proliferation depends heavily on features of individual technologies. With many weapons that we seek to limit access to, such as nuclear weapons, proliferation is not stopped by restricting knowledge (which is typically very difficult) but by restricting access to materials like enriched uranium. It seems like there is no significant materials shortage for BCI technology, as current BCIs do not require any fundamentally rare or unique materials. Furthermore, it is easier to prevent proliferation of technologies used by governments and militaries than it is to prevent proliferation of technologies used by civilians. Because BCIs are currently being planned as a widespread commercial technology, available to civilians, other countries are likely to gain access to them and have the opportunity to reverse engineer them. With this in mind, anti-proliferation methods would need to focus on preventing spread of knowledge about the development or security of BCIs – an incredibly difficult task for consumer products.

The possibility of defence

What future efforts will be available to counteract invasive surveillance and control of people's minds? Possible future countermeasures to address this risk factor are hard to accurately predict, as is their level of effectiveness, but this section will illustrate a few potential options.

One possible approach is to prevent the creation and widespread use of BCIs. This could be done culturally, through stigmatisation of BCIs used for non-medical purposes, or this could be done

through policy. Domestic legislation could ban the development or use of augmentative BCIs, or could introduce limits around data collection. International laws could assist in creating norms against the use of non-medical BCIs, as has been done with international conventions such as the Ottawa Treaty banning land mines and the Chemical Weapons Convention. Theoretically, similar measures could make widespread use of BCIs less likely and could decrease risk from BCIs overall.

However, the precedent of success here is mixed. Yonck (2020) argues that pushes for laws around data collection can be expected (based on the handling of current privacy struggles), but are unlikely to pass. Policy has had very little success addressing the expansion of government surveillance in the past, and it is likely that laws will continue to lag behind. However, stigmatisation (both through cultural and legal means) may be more tractable.

Of course, it is possible that individual countries may successfully legislate against (or build norms against) technology of this kind. However, there is also the issue of permanence. In times of crisis, many countries have rolled back political protections that proved inconvenient in a time of crisis. With BCIs, the same situation is likely. Even if countries are able to introduce legislation that restricts privacy, such legislation may collapse when it becomes inconvenient. This is true not just for laws, but also norms.

A second approach would be to attempt to guide the development of BCIs such that the technology develops in less dangerous ways. For example, actors could work to encourage the development of high fidelity, noninvasive, read-focused consumer BCIs. If consumers could be satisfied by the performance of these BCIs then it may reduce the demand to develop invasive technology. On the other hand, this strategy may not reduce demand at all. In showing what is possible this strategy may even drive demand for more invasive BCIs, and in doing so hasten the development of X-risk relevant BCIs. Guidance of this kind may be quite an unreliable approach.

A third avenue is to address BCIs not through politics, but through technological obsolescence. If there are technological advances that could make BCI reinforced control unsustainable or untenable for totalitarian governments, then the risk to humanity's future from BCIs could be lessened. I will unpack two possible options here, though there may be more.

a) As electronic devices, BCIs are currently vulnerable to EMPs (electromagnetic pulses). Weaponised EMPs are able to indiscriminately shut down electronics. An EMP could destroy all neural laces within a vicinity and liberate enslaved users from coercively implanted devices, reverting them to their uninfluenced state. This vulnerability to EMP attacks may also be a necessary weakness of BCI technology. This is because EMP shielding (at present) requires that objects be stored in a Faraday cage, which is able to block out electromagnetic radiation. This has two flaws. First, Faraday cages are very space intense; they are not a wearable device and could not be used at all times. Secondly, because Faraday cages block electromagnetic radiation BCIs would be unable to connect with other devices while in a Faraday cage – much like a phone that cannot get signal. This would significantly diminish the utility of BCIs. As such, BCIs will likely be vulnerable to EMPs for the foreseeable future.

b) Advances in cybersecurity may also be relevant to addressing risk from BCIs. If BCIs cannot be adequately defended, cyberattacks may work as a useful tool to liberate minds. Cyberattacks are often used maliciously but could also be used to shut down BCIs that are being used by totalitarian actors. If vulnerabilities were large enough, this could make widespread societal control through BCI use less viable. The exact vulnerabilities that could be exploited are hard to predict, as the technology is still in an early stage, and such prediction is beyond the scope of this paper. But it seems likely that cyberattacks could be a relevant tool to address the risk posed from BCIs.

At present, successful methods of defence against BCI reinforced totalitarianism are unclear. More research is urgently required.

Recommendations for Future Research

Given the insights from this paper, we recommend a few directions for future research.

1. More in depth analysis of the level of increase in risk caused by BCIs. In particular this would be assisted by stronger estimates on the baseline likelihood of totalitarian risk over the next 100 years.
2. A search for possible solutions that might reduce the level of risk caused by BCIs, or that might prevent the development of this risk factor.
3. Analysis of these solutions in terms of cost effectiveness.

Conclusion

This paper has sought to explore the potential of BCIs to increase the likelihood of long term global totalitarianism. We suggest that BCIs will allow for an unmatched expansion in the state's capacity for surveillance and will make it possible for states to police even the thoughts of their subjects. Secondly, brain stimulation will make both identifying and punishing dissent far more effective, and could potentially make dissenting thought a physical impossibility. We identify that with the development of BCIs totalitarianism would be far more internally stable, far more likely to last indefinitely, and would no longer require global spread to be sustainable in the long term. Due to these advantages, we argue that BCIs will offer strong incentives for dictatorships to adopt more totalitarian means of control, and thus make descent to totalitarianism more likely to occur.

Finally, we establish that BCIs set up an unusual strategic environment, where the existential risk is likely to become harder to solve over longer time periods. This gives further reason to address this risk sooner rather than later and put significant effort into either preventing the development of BCIs, or guiding their development in a safe way, if this is possible. Due to the current lack of discussion about this technology, and the high level of risk it poses, we believe that this risk factor deserves far more attention than it currently receives.

References

Abelson, J., Curtis, G., Sagher, O., Albucher, R., Harrigan, M., Taylor, S., Martis, B., & Giordani, B. (2005). Deep brain stimulation for refractory obsessive compulsive disorder. *Biological Psychiatry*, 57(5), 510-516. https://doi.org/10.1016/j.biopsych.2004.11.042

Allison, B., Wolpaw, E., & Wolpaw, J. (2007). Brain-computer interface systems: progress and prospects. *Expert Review of Medical Devices*, 4(4), 463-474. https://doi.org/10.1586/17434440.4.4.463

Anupama, H., Cauvery, N., & Lingaraju, G. (2012). Brain computer interface and its types - a study. *International Journal of Advances in Engineering and Technology*, 3(2), 739-745.

Bellman, C., Martin, M., MacDonald, S., Alomari, R., & Liscano, R. (2018). Have we met before? Using consumer-grade brain-computer interfaces to detect unaware facial recognition. *Computers in Entertainment*, 16(2), 7. https://doi.org/10.1145/3180661

Bernal, S., Celdran, A., Perez, G., Barros, M & Balasubramaniam, S. (2019a) Cybersecurity in brain computer interfaces: state-of-the-art, opportunities, and future challenges. https://arxiv.org/pdf/1908.03536.pdf.

Bernal, S., Huertas, A., & Perez, G. (2019b) Cybersecurity on brain-computer-interfaces: attacks and countermeasures. *Conference: V Jornadas Nacionales de Investigación en Ciberseguridad.*

Bittar, R., Kar-Purkayastha, I., Owen, S., Bear, R., Green, A., Wang, S., & Aziz, T. (2005a). Deep brain stimulation for pain relief: a meta-analysis. *Journal of Clinical Neuroscience*, 12(5), 515-519. https://doi.org/10.1016/j.jocn.2004.10.005

Bittar, R., Otero, S., Carter, H., & Aziz, T. (2005b). Deep brain stimulation for phantom limb pain. *Journal of Clinical Neuroscience*, 12(4), 399-404. https://doi.org/10.1016/j.jocn.2004.07.013

Bostrom, N. (2013). Existential risk prevention as global priority. *Global Policy*, 4(1), 15-31. https://doi.org/10.1111/1758-5899.12002

Bunce, S., Devaraj, A., Izzetoglu, M., Onaral, B., & Pourrezaei, K. (2005). Detecting deception in the brain: a functional near-infrared spectroscopy study of neural correlates of intentional deception. Proc. *SPIE Nondestructive Detection and Measurement for Homeland Security III*, 5769. https://doi.org/10.1117/12.600601

Burwell, S., Sample, M. & Racine, E. (2017). Ethical aspects of brain computer interfaces: a scoping review. *BMC Medical Ethics*, 18, 60. https://doi.org/10.1186/s12910-017-0220-y

Caplan, B. (2008). The totalitarian threat. In Nick Bostrom & Milan Cirkovic (eds) *Global catastrophic risks* (pp504-520). Oxford University Press.

Carmena, J., Lebedev, M., Crist, R., O'Doherty, J. Santucci, D., Dimitrov, D., Patil, P., Henriquez, C., & Nicolelis, M. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLOS Biology*, 1, 193-208. https://doi.org/10.1371/journal.pbio.0000042

Chenoweth, E. (2011). *Why civil resistance works*. Colombia University Press.

Constine, J. (2017). Facebook is building brain computer interfaces for typing and skin-hearing. https://techcrunch.com/2017/04/19/facebook-brain-interface/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAF-G70vjq6iq3xbyTrRJX142g75UeLop2lh4vADsXSRxkbukky53hw2ztInvOKfgHxB0fl1YRSQwKJWVEeeXBw8ArIJuNLVh0z2Qb9Pe7sBKUGiEY-a0jkh5PHcArqoPjIc_4srTbiBZzjwN7QsNV3CuooTODW9-TxsZnp5q9RBf.

Cotton-Barratt, O., & Ord, T. (2015). Existential Risk and Existential Hope. *Future of Humanity Institute - Technical Report*, 1.

DARPA. (2019). Six paths to the nonsurgical future of brain-machine interfaces. https://www.darpa.mil/news-events/2019-05-20.

Delgado, J.M.R. (1969). *Physical control of the mind: toward a psychocivilized society*. Harper and Rowe.

Demetriades, A.K., Demetriades, C.K., Watts, & K. Ashkan. (2010). Brain-machine interface: the challenge of neuroethics. *The Surgeon*, 8(5), 267–269. https://doi.org/10.1016/j.surge.2010.05.006

Deuschl, G., Schade-Brittinger, C., Krack, P., Volkmann, J., Schafer, H., Botzel, K., Daniels, C., Deutschlander, A., & Dillmann, U., et al. (2005). A randomized trial of deep-brain stimulation for parkinsons disease. *New England Journal of Medicine*, 355, 896-908. https://doi.org/10.1056/NEJMoa060281

Evers, K., & Sigman, M. (2013). Possibilities and limits of mind-reading: a neurophilosophical perspective. *Consciousness and Cognition*, 22, 887-897. https://doi.org/10.1016/j.concog.2013.05.011

Gaul, V. (2020). Brain computer interface market by type (invasive BCI, non-invasive BCI and partially invasive BCI), application (communication & control, healthcare, smart home control, entertainment & gaming, and others): global opportunity analysis and industry forecast, 2020-2027. Allied Market Research. https://www.alliedmarketresearch.com/brain-computer-interfaces-market.

Glannon, W. (2009). Stimulating brains, altering minds. *Journal of Medical Ethics*, 35, 289–292. http://dx.doi.org/10.1136/jme.2008.027789

Greenberg, B., Malone, D., Friehs, G., Rezai, A., Kubu, C., Malloy, P., Salloway, S., Okun, M., Goodman, W., & Rasmussen, S. (2006). Three-year outcomes in deep brain stimulation for highly resistant obsessive–compulsive disorder. *Neuropsychopharmacology*, 31, 2384–2393. https://doi.org/10.1038/sj.npp.1301165

Guenther F.H., Brumberg, J.S., Wright, E.J., Nieto-Castanon, A., Tourville, J.A., Panko, M., Law, R., Siebert, S.A., Bartels, J., Andreasen, D., Ehirim, P., Mao, H., & Kennedy, P. (2009). A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE*, 4(12), e8218. https://doi.org/10.1371/journal.pone.0008218

Gulati, T., Won, S.J., Ramanathan, DS., Wong, C., Bopepudi, A., Swanson, R., & Ganguly, K. (2015). Robust neuroprosthetic control from the stroke perilesional cortex. *Journal of Neuroscience*, 35(22), 8653-8661. https://doi.org/10.1523/JNEUROSCI.5007-14.2015

Halpern, C.H., Wolf, J.A., Bale, T.L., Stunkard, A.J., Danish, S.F., Grossman, M., Jaggi, J., Grady, S., & Baltuch, G. (2008). Deep brain stimulation in the treatment of obesity. *Journal of Neurosurgery*, 109(4), 625–634. https://doi.org/10.3171/JNS/2008/109/10/0625

Hamani, C., McAndrews, M., Cohn, M., Oh, M., Zumsteg, D., Shapiro, C., Wennberg, R., & Lozano, A. (2008). Memory enhancement induced by hypothalamic/fornix deep brain stimulation. *Annals of Neurology*, 63, 119-123. https://doi.org/10.1002/ana.21295

Ienca, M. (2015). Neuroprivacy, neurosecurity and brain hacking: emerging issues in neural engineering. *Bioethica Forum*, 8(2), 51-53.

Ienca, M., & Haselager P. (2016). Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18, 117-129. https://doi.org/10.1007/s10676-016-9398-9

Ifft, P., Shokur, S., Li, Z., Lebedev, M., & Nicolelis, M. (2013). A brain machine interface that enables bimanual arm movement in monkeys. *Science Translational Medicine*, 5(210), 210ra154. https://doi.org/10.1126/scitranslmed.3006159

Kernel: neuroscience as a service. (2020). https://www.kernel.co/.

Klein E., Brown T., Sample M., & Truitt AR, Goering S. (2015). Engineering the brain: ethical issues and the introduction of neural devices. *Hastings Center Report*, 45(6), 26–35. https://doi.org/10.1002/hast.515

Kotchetkov, I. S., Hwang, B.Y., Appelboom, G., Kellner, C.P., & Connolly, E.S. (2010). Brain-computer interfaces: military, neurosurgical, and ethical perspective. *Neurosurgical Focus*, 28(5). https://doi.org/10.3171/2010.2.FOCUS1027

Kumar, K., Toth, C., & Nath, R.K. (1997). Deep brain stimulation for intractable pain: a 15-year experience. *Neurosurgery*, 40(4), 736-747. https://doi.org/10.1097/00006123-199704000-00015

Lammel, S., Lim, B.K., Ran, C., Huang, K.W., Betley, M., Tye, K., Deisseroth, K., & Malenka, R.(2012). Input-specific control of reward and aversion in the ventral tegmental area. *Nature*, 491, 212–217. https://doi.org/10.1038/nature11527

Laxton, A.L., & Lozano, A.M. (2013). Deep brain stimulation for the treatment of Alzheimer's disease and dementias. *World Neurosurgery*, 80(3-4), 28.el-S28.e8. https://doi.org/10.1016/j.wneu.2012.06.028

Mayberg, H., Lozano, A., Voon, V., McNeely, H., Seminowicz, D., Hamani, C., Schwalb, & J., Kennedy, S. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, 45(5), 651-660. https://doi.org/10.1016/j.neuron.2005.02.014

Mazzoleni, M., & Previdi, F. (2015). A comparison of classification algorithms for brain computer interface in drug craving treatment. *IFAC Papers Online*, 48(20), 487-492. https://doi.org/10.1016/j.ifacol.2015.10.188

Merrill, N., & Chuang, J. (2018). From scanning brains to reading minds: talking to engineers about brain computer interface. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Paper no 323. 1-11. https://doi.org/10.1145/3173574.3173897

Moore, M.M. (2003). Real-world applications for brain-computer interface technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2), 162-165. https://doi.org/10.1109/TNSRE.2003.814433

Moses, D.A., Leonard, M.K., Makin, J.G., & Chang, E. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, 10, 3096. https://doi.org/10.1038/s41467-019-10994-4

Munyon, C. (2018). Neuroethics of non-primary brain computer interface: focus on potential military applications. *Frontiers in Neuroscience*, 12, 696. https://doi.org/10.3389/fnins.2018.00696

Musk, E. (2019). An integrated brain-machine interface platform with thousands of channels. *Journal of Medical Internet Research*, 21(10). https://doi.org/10.2196/16194

Nijboer, F., Clausen, J., Alison, B., & Haselager, P. (2011). The Asilomar survey: stakeholders opinions on ethical issues related to brain-computer interfacing. *Neuroethics*, 6, 541-578. https://doi.org/10.1007/s12152-011-9132-6

Ord, T. (2020). *The Precipice.* Oxford University Press.

Orwell, G. (1945). You and the atomic bomb. Tribune (19 October 1945), reprinted in *The collected essays: journalism and letters of George Orwell, vol. 4: in front of your nose* (pp6-9). 1945–1950, ed. Orwell. S. and Angus. I.(Seeker and Warburg, 1968). www.orwell.ru/library/articles/ABomb/english/e_abomb

Perlmutter, J., & Mink, J. (2006). Deep brain stimulation. *Annual Review of Neuroscience*, 29. 229-257. https://doi.org/10.1146/annurev.neuro.29.051605.112824

Roelfsema, P., Denys, D & Klink, P.C. (2018). Mind reading and writing: the future of neurotechnology. *Trends in Cognitive Sciences*, 22(7). 598-610. https://doi.org/10.1016/j.tics.2018.04.001

Sharp, G. (1973). *The politics of nonviolent action.* MA: Porter Sargent.

Statt, N. (2017). Kernel is trying to hack the human brain - but neuroscience has a long way to go. https://www.theverge.com/2017/2/22/14631122/kernel-neuroscience-bryan-johnson-human-intelligence-ai-startup).

Suthana, N., Haneef, Z., Stern, J., Mukamel, R., Behnke, E., Knowlton, B., & Fried, I. (2012). Memory enhancement and deep-brain stimulation of the entorhinal area. *New England Journal of Medicine*, 366, 502-510. https://doi.org/10.1056/NEJMoa1107212

Torres, P. (2016). Climate change is the most urgent existential risk. Future of Life Institute. https://futureoflife.org/2016/07/22/climate-change-is-the-most-urgent-existential-risk/?cn-reloaded=1.

Tsai, H.C., Zhang, F., Adamantidis, A., Stubler, G., Bonci, A., Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, 324(5930), 1080–1084. https://doi.org/10.1126/science.1168878

Tucker, P. (2018). Defence intel chief worried about Chinese 'integration of human and machines' in defence one. https://www.defenseone.com/technology/2018/10/defense-intel-chief-worried-about-chinese-integration-human-and-machines/151904/.

Wu, S., Xu, X., Shu, L., & Hu, B. (2017). Estimation of valence of emotion using two frontal EEG channels. *2017 IEEE international conference on bioinformatics and biomedicine* (pp. 1127–1130). https://doi.org/10.1109/BIBM.2017.8217815

Yonck, R. (2020). *Future minds: the rise of intelligence from the big bang to the end of the universe.* Simon and Schuster.