Wireheading as a Possible Contributor to Civilizational Decline

August 2018

*Alexey Turchin*

Digital Immortality Now

Foundation Science for Life Extension

alexeiturchin@gmail.com

*David Denkenberger, PhD*

University of Alaska Fairbanks

Global Catastrophic Risk Institute

Alliance to Feed the Earth in Disasters (ALLFED)

Abstract: Advances in new technologies create new ways to stimulate the pleasure center of the human brain via new chemicals, direct application of electricity, electromagnetic fields, “reward hacking” in games and social networks, and in the future, possibly via genetic manipulation, nanorobots and AI systems. This may have two consequences: a) human life may become more interesting, b) humans may stop participating in any external activities, including work, maintenance, reproduction, and even caring for their own health, which could slowly contribute to a decline in human populations. As wireheading is not self-limiting, it may push the addict to his/her limits and ultimately result in life-threatening behavior. Mass unemployment and the provision of basic income would likely feed the industry of pleasures. DIY synthetic biology and computers may create the possibility of homemade very addictive drugs through various routes, like insertion of a morphine-producing gene in plants or stimulation of the brain via electromagnetic fields, which may not be possible to control via limiting the supply chain. If “wireheading knowledge” becomes freely available, it could create a reward-hacking epidemic larger than the current U.S. opioid crisis and possibly contribute to a global civilizational decline.

**Key points:**

* Technological progress increases the number of ways, intensity and availability of the ways of brain stimulation.
* DIY brain stimulation methods could be disseminated as pure knowledge.
* Unlimited human reward center stimulation could cause death or significant impairments in many ways.
* Reward center stimulation technology alone is very unlikely to cause human extinction directly.
* Reward center stimulation technology could contribute to possible civilizational collapse via lowering fertility, overdose mortality, mortality from lack of self-care, etc, and this could eventually result in extinction.

[1 Introduction 3](#_Toc522031780)

[2 Historical examples of societal problems arising from mass addictions 4](#_Toc522031781)

[2.1 Connection between different ways of reward hacking and social problems 4](#_Toc522031782)

[2.2 Historical growth of the number and effectiveness of methods for the reward center stimulation 5](#_Toc522031783)

[3 New ways of stimulating the pleasure center in the brain 5](#_Toc522031784)

[3.1 Overview of the human reward system and its electrical stimulation 6](#_Toc522031785)

[3.2 Different ways to reach human reward center via new technologies 6](#_Toc522031786)

[3.3 Use of electric fields and currents for brain stimulation 7](#_Toc522031787)

[3.4 Neuroimplants and other invasive methods 8](#_Toc522031788)

[3.5 Brain stimulation and AI 9](#_Toc522031789)

[4 Possible catastrophic consequences 10](#_Toc522031790)

[4.1 The ways how reward stimulation and addiction could cause death 10](#_Toc522031791)

[4.2 Catastrophic wireheading criteria 12](#_Toc522031792)

[4.3 Positive world of eternal pleasures 13](#_Toc522031793)

[5 Wireheading in the context of other catastrophic risks and civilizational problems 14](#_Toc522031794)

[5.1 Relation between drug legalization, technological unemployment and basic income 14](#_Toc522031795)

[5.2 Could widespread wireheading be explanation of the Fermi paradox? 15](#_Toc522031796)

[5.3 Possible protective measures 15](#_Toc522031797)

[Conclusion 16](#_Toc522031798)

[Literature 16](#_Toc522031799)

*God, give me a chance not to stumble,*

*Not to drown in this love:*

*We still have to live, fly to the stars,*

*Learn the secrets of the vast universe!*

*Trillions of souls are outside the window*

*Their life is only a moment, sad and sleepy.*

*They cannot achieve happiness by themselves*

*Who else but us can give it to them?*

V. Argonov, techno-opera

“Rethinking progress: beyond the horizon” (Argonov 2015)

# 1 Introduction

Pleasure is the reward function created by powerful evolutionary pressure to motivate human minds do what is dictated by innate drives, like sex, survival, eating and even growth of social status (Simler and Hanson 2017). This reward function is effectively protected from self-hacking in animals living in nature, perhaps because species capable of such self-stimulation become extinct. Human civilization created many new ways to stimulate pleasure centers, including art, protected sex, partying, etc. However, some of the ways to hack the reward function are so effective that they effectively incapacitate the individual and create persistent social problems (e.g. alcoholism). Moreover, as pleasure seeking is one of the main motives of human behavior, humans are very creative in inventing new ways to stimulate pleasure centers.

Here we investigate a new global risk which has been barely mentioned in the academic literature, with some forays in fiction. Advances in new technologies create the possibility of affecting reward centers of the brains in new ways which could be even stronger and could be disseminated as pure knowledge. For example, the idea that a cheap, super-addictive homemade drug could appear was explored in the Strugatskys’ novel “The final circle of paradise” (Strugatsky and Strugatsky 1976). It is described as an electronic drug which creates such pleasant and realistic hallucinations that users do not want to return to reality and die of exhaustion; the drug is very simple and could be built from an ordinary radio. The same problem is addressed in Argonov’s techno-opera “Beyond the line”, which described crystals of artificial “love” (similar to 3,4-Methyl​enedioxy​methamphetamine, MDMA) which could seduce humanity to stop space exploration and instead choose infinite pleasures now (Argonov 2015) and in the StarTrek episode “The Game” (Allen 1991).

The term “wireheading” appeared as a short description of the deep brain stimulation in rat experiments (Olds and Milner 1954). The rats were able to self-stimulate by pressing a pedal and they eventually died from exhaustion (in many cases, the experiments were stopped before they died for anatomic study, and we will look later at how different causes of death are distributed for drugs self-administrated by animals).

While wireheading is often used as a scary tale, many utilitarian philosophers have written positively about it, and the risks of global catastrophic it poses is underexplored. J. Taylor started her article about wireheading with a quote from Dalai Lama about electrodes: “If it was possible to become free of negative emotions by a riskless implementation of an electrode – without impairing intelligence and the critical mind – I would be the first patient” (Taylor 2012). She then argues for cheap and safe wireheading; she claims it will increase compassion and motivation and cites experiments where depressed people self-stimulated only when they were depressed. Similarly, Tomasic wrote that humanity as whole is moving in the direction of global wireheading (Tomasik 2015). The whole site *wireheading.com* is devoted to possible “positive wireheading” in the form of paradise engineering.

Only a few accidental human wireheading cases are known to date; however, they are not examples of “full-blown” reward hacking, as only some part of reward centers was stimulated (Portenoy et al. 1986). Wireheading is a problem not only for humans, but also for AIs which use reward functions as was described by Yamploskiy (2014). Experiments on rats have shown that a rat will constantly press a lever to get the next drip of cocaine until the point of starvation and even death (Olds and Milner 1954).

Some scientists have looked at past reward-hacking technologies as a cause of societal crises (Khalturina and Korotaev 2008). In this article we explore if the trend to more and more effective human wireheading via quick advances of several technologies could result in civilizational scale consequences, like social collapse, global catastrophe, or even human extinction.

In section 2 we provide an overview of known examples of catastrophic wireheading and the current methods to stimulate the human brain. In section 3 we look more deeply into the nature of the human reward center and those technologies which currently exist or could appear for its stimulation. Section 4 builds a bridge between wireheading and global catastrophic risks by exploring the ways in which wireheading could kill and spread to become a global problem. In section 5, the global risk of wireheading epidemics are placed in the wider perspective of accelerating technological progress, technological unemployment, and other catastrophic risks.

# 2 Historical examples of societal problems arising from mass addictions

## 2.1 Connection between different ways of reward hacking and social problems

Addictions have created persistent social problems throughout history but have never reached a globally catastrophic level.

**Alcohol addiction** is known to be a widespread social problem in many countries. In Russia, it was shown to be responsible for super-mortality in the second half of the 20th century (Khalturina and Korotaev 2008). Alcohol hit people of eastern Asian origin and native people, who do not have an evolutionary adaption to alcohol in the form of enzymes, especially hard. High alcohol consumption (above 100 g a week) has reduced the average life expectancy by five years, according to recent meta-analysis (Wood et al. 2018).

Alcohol addiction is slow but can kill in several ways: direct damage to health, crimes committed under the effect of alcohol, alcohol-related accidents, and fetal damage during pregnancy. The most devastating effects resulted from the introduction of alcohol to Siberian and American native populations, which suffered a significant decline. But even in Russia, alcoholism is currently in decline as people find better alternatives like social nets and games (Neyaskin 2018).

**Smoking** has significantly contributed to shorter life expectancy in the contemporary world. Some estimates suggest that it reduces life expectancy by an average of 10 years (CDC 2018). Nicotine is extremely addictive, as it potentiates learning, and the brain learns to associate pleasure with cigarettes. Nicotine itself is not as toxic as some other chemicals in tobacco smoke.

**Overeating** and the corresponding obesity epidemic is also connected with reward self-stimulation via eating foods high in sugar and fat.

**Opiate addiction** contributed to a social crisis in 19th century China. Currently in the U.S., 49 000 people in 2017 died because of opiate overdoses and 72 000 from all other drugs (National Institute on Drug Abuse 2018), the other drugs making up around 2.6 per cent of total mortality of 2 712 630 in 2015 (National Center for Health Statistics 2018). The current opioid crisis in the U.S. represents an exponential growth in the number of overdose deaths due to heroin and synthetic opioids, from thousands in the 2000s to tens of thousands in the 2010s. Its causes are complex, but one cause is connected with technological advances in the production of more and more effective opioid receptor stimulators. The illicit dilution of heroin by the more-potent fentanyl has increased the probability of overdoses.

**HIV** takes advantage of human addiction for transmission: addiction to heroin (via injections) and some would argue to sex. In South Africa the HIV epidemic reduced life expectancy at peak for 10 years, but the situation recently improved with anti-retrovirus therapy (Suthar and Bärnighausen 2017).

**Gaming addiction:** some people have played computer games for days while ignoring their bodily needs and died (Reuters 2007). Others become addicted to gambling. Current gaming cannot be catastrophic as millions of people play but only a few have negative consequences like worsening health and obesity because of sedimentary life-style and de-socialization. The important point here is that a carefully designed reward may not need a physical brain intervention to create intense pleasure, perhaps if it creates some mental program, typically around a numerical score.

**Social network addiction** is based around scoring feedback in the form of “likes” (Andreassen 2015). Some people have undertaken dangerous activity to get great shots for Instagram and are killed, e.g. by falling from heights or being mauled or trampled by animals (Polianskaia 2018). Other people who get a small number of likes become depressed and need to compensate by other pleasure-seeking activity or risk suicide (Oberst et al. 2017).

**Extreme sport and dangerous activity**. Some people stimulate their pleasure centers by deliberately taking risks. Examples include risky driving, extreme sports, participation in criminal activity, and even roller coasters. Unsafe driving is probably the deadliest of the “adrenaline” addictions.

**Hacking social reward as crime**. Most criminal activities are attempts to hijack the socially accepted reward function: stealing is about money, which is a reward in a correctly functioning economy, rape is about hijacking consensual sex, and drug-related crimes are obviously connected to either money or hijacking of the reward function.

Some addictions may be beneficial (Miller 2018), like those to sport, career, or making money. However, any addiction tends to push the addict to a limit at which it becomes detrimental because of the growing tolerance or slippery slope to reach higher pleasures.

In general, society is able to cope with dangerous addictions, as they either progress slowly, and so allow time for countermeasures to be developed, or they affect only a small subpopulation, the collapse of which is not damaging to civilization as a whole.

## 2.2 Historical growth of the number and effectiveness of methods for the reward center stimulation

During human history, there has been a slow growth of methods to hack the human reward function. In ancient times, only a few ways to get pleasures without fulfilling biological functions were known, including alcohol, opium, overeating, non-reproductive sex, and gambling. However, it is interesting to note that biologically useless pleasures become important pillars of the social structures, like collective drinking, and can support the economy (e.g. wine production). The discovery of the New World allowed exchange not only of infectious diseases but also of new ways to get pleasure, the advent of nicotine and cocaine in the West.

Advances in chemistry in the 20th century brought us heroin, LSD, and many new synthetic drugs, including antidepressants. Brain research produced electric wireheading. At the beginning of the 21st century, thousands of psychoactive drugs were known and it became possible to predict—to some extent—the properties of new chemicals, which created the phenomenon of “designer drugs”. In this phenomenon, known drugs are changed to escape legal barriers or for cheaper manufacturing. More potent drugs provide more income to drug cartels as they earn more money from each kilogram of manufactured drug; smaller quantities are also simpler to illegally transport (Frank and Pollack 2017).

In general, the law of techno-humanitarian balance (Nazaretian 2004) is applicable not only to military production, but to new “reward hacking” ways to get pleasures: society is able to adapt either by banning these new instruments or by including lighter forms of them within the social structure. But will it be possible to continue to support such a balance in a situation of very rapid technological progress? Anything we create to interact with the brain or to improve mood and motivation could be converted into a reward stimulation system.

# 3 New ways of stimulating the pleasure center in the brain

*We had two bags of grass, seventy-five pellets of mescaline, five sheets of high powered blotter acid, a salt shaker half full of cocaine, and a whole galaxy of multi-colored uppers, downers, screamers, laughers... and also a quart of tequila, a quart of rum, a case of Budweiser, a pint of raw ether and two dozen amyls.
Not that we needed all that for the trip, but once you get locked into a serious drug collection, the tendency is to push it as far as you can.*

“Fear and Loathing in Las Vegas”, (Gilliam 1998)

## 3.1 Overview of the human reward system and its electrical stimulation

The human reward system consists of several interconnected nodes. If one node is stimulated, others fire too (Yager et al. 2015). Intracranial self-stimulation in rats is possible via electrodes implanted in the medial forebrain bundle (Carlezon Jr and Chartoff 2007), as it requires less current than other parts of reward system.

It was found that humans have two interconnected systems, roughly corresponding to “wanting” and “liking”. The fact of heroin addiction in rats and humans is evidence that electric self-stimulation of the pleasure center is also possible in humans.

One human case study draws a bleak picture of a woman who became addicted to electrical self-stimulation, which was not a full reward stimulation, but only partial orgasmic stimulation. She had intractable pain in her vertebrae and so had several deep brain-stimulating electrodes inserted in her brain.

Soon after insertion of the nVPL electrode, the patient noted that stimulation also produced erotic sensations. This pleasurable response was heightened by continuous stimulation at 75% maximal amplitude, frequently augmented by short bursts at maximal amplitude. Though sexual arousal was prominent, no orgasm occurred with these brief increases in stimulation intensity. Despite several episodes of paroxysmal atria tachycardia and the development of adverse behavioral and neurological symptoms during maximal stimulation, compulsive use of the stimulator developed. At its most frequent, the patient self-stimulated throughout the day, neglecting personal hygiene and family commitments. A chronic ulceration developed at the tip of the finger used to adjust the amplitude dial and she frequently tampered with the device in an effort to increase the stimulation amplitude. At times, she implored her family to limit her access to the stimulator, each time demanding its return after a short hiatus. During the past 2 years, compulsive use has become associated with frequent attacks of anxiety, depersonalization, periods of psychogenic polydipsia, and virtually complete inactivity (Portenoy et al. 1986).

Deep brain stimulation (DBS) via implanted electrodes deep into human brain was approved by the FDA in 2003 for Parkinson disease, but not for depression, which is still under clinical trials. For depression, DBS targets exactly the medial forebrain bundle which is involved in the reward and motivation, and it is not surprising that affecting regions close to reward center improves mood (Schlaepfer et al. 2013). DBS is major surgery; its price and complexity is prohibitive for recreational use, with costs in the range of 30 000–50 000 USD (Goodman 2011).

## 3.2 Different ways to reach human reward center via new technologies

The number of methods for human wireheading is growing exponentially, and at the same time, the price is declining, thanks to advances in neurobiology, synthetic biology and psychology. In this section, we analyze methods of pleasure center stimulation which could appear in the 21st century.

**Genetically modified plants** and bacteria could produce addictive and psychotropic drugs. Many drugs are relatively simple molecules which are synthesized by a cell, like cocaine, psilocybin, and opiates, and are encoded by just a single gene or group of genes (Docimo et al. 2012). Such a gene could be inserted in other plants or bacteria, and such technology is already used for synthesis of many medical drugs, like insulin (Baeshen et al. 2014). In another example, a hydrocodone-producing yeast was created in 2015 (Galanie et al. 2015).

This approach offers the opportunity for criminals to hide their activity, as genetically modified plants or yeast will look like ordinary strains of the species. However, if such modified plants leak into environment, they could be a significant health hazard, as an animal or a person could eat them and overdose. People will likely grow such plants at home, in the garden or in the forest nearby, escaping the need to pay dealers for drugs. It is possible that they could continually redose and kill themselves in the process.

Advances in the understanding of brain chemistry and architecture, as well as in biosynthesis and drug discovery, will open the possibility of new designer drugs with different properties. But humans react to drugs individually, and some could die from the first dose of cocaine (Browne 2011), so starting self-experimentation with different drugs may end in an eventual individual adverse reaction.

“**Orgasmotron**”. One may not even need reach deep into a brain to get pleasure center stimulation. A device built to cure back pain accidentally created orgasms (ABC 2017). Sexual self-stimulation with asphyxia (and the “choking game” among teenagers (Linkletter et al. 2010)) is an example of a “knowledge-only” dangerous addictive drug. This killed at least 82 people in US in 1995-2007 (Toblin et al. 2008).

**Physiological stimulation**. Warm water, and some forms of physical activity, like the so-called “runner’s high” and kundalini yoga could increase reward center stimulation. Concentration of attention on some points of the body’s surface is known to change the practitioner’s psychological state and is used in some forms of Eastern meditation (Amihai and Kozhevnikov 2015).

**Drug combinations**. Some drug combinations are known to be especially dangerous, like the combination of cocaine and heroin colloquially known as “speedballs”. But more sinister is the situation when a combination of some benign things becomes addictive, like small money prizes and a random number generator combining to create a gaming addiction.

**Audio stimulation** by relatively conventional methods, for example, in the form of music or drums, is part of human cultural heritage. New ideas have appeared in the field, like binaural rhythms (Oster 1973), isochronic tones, brain machines using pulsing light (e.g. “Kasina”), but most of them produce only limited and temporal mind-changing effect. Biofeedback also may be used to train the brain to stimulate its reward centers.

## 3.3 Use of electric fields and currents for brain stimulation

Electromagnetic fields also offer opportunities for brain stimulation. There are several new ways to non-invasively affect brain states, including transcranial direct and alternating current stimulation, transcranial magnetic stimulation, and infrared and ultrasound stimulation. The regions for reward center stimulation are deep inside the brain and thus not easily reachable. Regardless, the market for neurostimulation devices is growing and is expected to reach 13 billion USD in 2023 (Global Market Insights 2017).

While external brain stimulation is not precise enough to target only the reward center, the use of the mathematical methods, brain imaging and AI may provide much better targeting in the future. The complex activity of many electromagnetic coils could be used to create a field of complex-space geometry to stimulate precisely the desired region of the brain using Fourier analysis to get the needed form of the field. Pierre and Persinger have created a helmet which reportedly induces religious visions; it uses some form of transcranial magnetic stimulation via complex set of coils which mimic the effects of temporal lobe epilepsy (Pierre and Persinger 2006).

**Lasers are used for brain stimulation** in the form of low-level light therapy (LLLT) (Disner et al. 2016). The intersection of many infrared lights from lasers could be able to reach some regions in the brain, as the tissues of the head including the skull are relatively transparent to infrared light.

**Transcranial direct current stimulation** (tDCS) devices could be as simple as 9V battery with two wires. We now know that they do not cause euphoria, but there were medical concerns raised at their initial unveiling. Anecdotal reports say that mood immediately improves when tDCS is used for therapy in patients with depression (Franson 2013).

**Cranial electrotherapy stimulation** (CES) is the use of alternating current to mimic some brain frequency to change working of the brain. There is less research on this approach than on tDCS. It has been suggested that it could be effectively used to induce lucid dreams. The Delight Pro is a consumer device which combines CES with audio-video stimulation for “entrainment” of brain waves (Delight Pro 2018).

**Pulsed magnetic field therapy** (PEMF) uses a changing magnetic field to stimulate areas deep inside the brain via induced currents. This approach is different from transcranial magnetic stimulation (TMS), which uses shorter but more intense pulses created by very powerful coils to suppress activity of a brain region (Groppa et al. 2012).

## 3.4 Neuroimplants and other invasive methods

It seems that the most powerful methods of stimulation are via direct electric stimulation of brain tissue, a fact which has given rise to the name “wireheading”. Chemical stimulation could lose its efficiency as tolerance is built. Electric stimulation does not have these limitations, as it directly stimulates neurons.

Fortunately, the human reward center is in a protected location deep within the brain, which is in turn protected by the skull. Must humans cannot reach it without complex invasive surgery. If a person is not yet addicted to brain stimulation, s/he generally will not risk brain surgery to obtain it. There are only a few examples of people who have tried voluntary brain surgery in non-medical contexts, mostly for experimental proposes, not for stimulation. For example, there are non-traditional medical procedures, like trepanation for brain enhancement (Cox 2013).

The advancing field of neuroimplants will make future brain surgeries much simpler, safer, and more standardized. It is still unlikely that people new to electric reward-center stimulation will seek this type of neuroimplants, as they will know that reward center constant self-stimulation could kill them in a few days. But if they were sure that they would have access to large supply of food and care, they could prefer constant wireheading bliss to, say, simple suicide.

However, the situation may be different with neuroimplants based on some form of nanobots or smart powder, where a person could control where they move in his/her brain. There would be a tendency to move the neuroimplant closer and closer to neural pathways connected with the reward center, with eventual wireheading. This can be considered an analogy to the way as a person may advance from light recreational drugs to heavy ones, or increase doses of analgesics.

**Electrode arrays.** DBS via a single electrode is a rather unsophisticated way to get pleasure, as such stimulation does not take into account the complex structure of the human reward center and stimulates the whole area. Thus—presumably—it cannot reach the maximum level of pleasure. However, a more powerful system might include many smaller electrodes in many parts of the brain connected to specific groups of neurons. This approach may be more difficult: science do not yet know the structures of these pleasure centers; smaller electrodes degrade more quickly, losing electric connections with the brain tissue or killing neurons around them; and such surgery is also currently difficult.

**Neural prosthesis.** The most common current implants are cochlear implants. They are connected to nerve tissue, not brain tissue, have up to 22 electrodes (Van Besouw 2013) and cost up to 100 000 USD. A more advanced neural prosthesis has been researched for the hippocampus (Gonzalez 2018) as memory improvement implants. In DBS, no direct current is used, but instead a combination of pulses of some frequency (like 100 Hz) which is regulated by a signal processor (Fagundes et al. 2016). For most effective reward stimulation, a special digital processor is needed which would create different signals for different parts of the brain. It could be put inside the brain together with electrodes, creating something like a reward-stimulating implant.

**Neural dust**. This idea is to create small self-sufficient electric machines which could be injected in the brain where they would exchange information and recharge wirelessly. The idea is explored in Musk’s project Neuralink (Templeton 2017).

**Targeted chemicals and nanoparticles.** The “advantage” of chemicals like opioids in reward center stimulation is that they non-invasively find their way into the brain via the blood stream, cross the blood–brain barrier and reach the target cells, which have unique receptors to which the molecules bind (Merrer et al. 2009). Specially designed ligands or nanoparticles functionalized with such ligands could be used in the same manner, to permanently attach to reward center cells and then be used as the basis of a system for its constant stimulation.

**Optogenetics and living neurons as brain implants**. Some viral vectors could be used to deliver genes (Ramos et al. 2017) that code for light-sensitive proteins into specific neurons, and such neurons could later be stimulated by sending infrared laser light in their direction.

**Artificial biological neurons as implants.** If humanity’s understanding of biological tissue becomes sufficiently advanced, we could harness the implanting mechanism used during the early embryonal development of the human brain. This is the ability of new neurons to travel through the brain, find a specific location, and from there, send axonal connections to other brain regions. Axons from live neurons could grow into remote regions of the brain and be used as relatively safe brain implants (Adewole et al. 2018). This technology would have important applications, like restoring brain function, creating non-degradable brain–computer interfaces, and ultimately, uploading brains, but it could be also used for wireheading.

**Nanorobots**. Classical self-replicating nanorobots (Drexler 1986) may be the ultimate brain hacking machines. Molecular manufacturing will also enable production of all possible drugs.

**Qualia engineering.** By using different neural implants, technology could not only stimulate the brain centers, but create new qualia of pleasures, more intense or more variable.

## 3.5 Brain stimulation and AI

It seems likely that in the future, AI will be the main player in the entertainment industry. AI will advance in computer games, virtual reality (VR), movies and production of other art forms. External computers with some form of AI will be used to control neuroimplants and nanobots in brains (Price 2018).

**AI optimizes reward in games and social networks**. An increase in the human reward response could be seen as an optimization problem, and AI could find individual ways to increase perceived reward. This might be in much the same way as Facebook’s algorithm increases people’s social network addiction by manipulating what they see and encouraging them to create more likable posts. This could be seen as an adaptive computer game, which perfectly stimulates one’s pleasure centers via interesting plot, neural interfaces and many other subtler clues, like sounds, which maximize addiction or reward.

**AI-empowered VR games** which also include brain stimulation could put a person in a complete simulated reality including meeting virtual sex partners or using imaginary recreational drugs.

**Drug-enhanced lucid dreams and dream recording** could harness the natural human ability to bypass the wall to the pleasure center while dreaming, using some combination of drugs, e.g. galantamine (Sparrow et al. 2018), lucidity-inducing glasses, and brain stimulation.

**Sexual AI-empowered “sexbots”** may develop a way to the human reward center via simulating something like multiple orgasm, or just destroy the human family and lower fertility. In some similar sense, owning a pet could satisfy a desire for children and thus contribute to lower fertility: the pet is reward-hacking for desire of taking care of something small.

**A global “Friendly” AI** with faulty programming could wirehead humans, thinking that orgasmium is what we actually want (Turchin and Denkenberger 2018a). The problem is not easily escapable, as if people had an AI system, many would want to ask it to make every given situation better, and eventually it may fall into a death spiral of asking for more and more pleasant experiences.

**Human uploads**, or ems (Hanson 2016), will have higher-level access to their reward function and thus be likely to slip into wireheading.

**Future global AI collapse**. A future possible global AI working as a world government (Singleton AI) could suddenly wirehead itself and halt, that is, find the way to bypass doing useful work and increase its own reward to maximum (Yampolskiy 2014, Turchin and Denkenberger 2018a). In that case, all robotic systems it controls will probably halt or start deviant behavior, including medical implants, home robots and transport.

**Future Suffering Risks (S-risks).** Knowledge of manipulating the human reward center creates the technology to produce intense suffering by stimulating the nearby centers of pain. It could be turned into a suffering mode either accidentally, if a Global AI has some kind of fatal error, like in the plot of (Ellison 1967), or with the goal of torture for blackmail. Turning off the stimulation in an addict will make the person crave stimulation and suffer from “wireheading withdrawal”. Last but not least, even constant wireheading could be regarded as a form of suffering from the point of view of the other goals (in preferential utilitarianism) or other parts of the brain, which could have qualia of suffering at the same time, like hunger.

# 4 Possible catastrophic consequences

## 4.1 Mechanisms of reward stimulation and addiction leading to death

There are many ways reward center stimulation could kill a person or damage society. Most of them are personally but not globally catastrophic. Additional conditions are needed to make an addiction a global danger.

Here we first look at the ways in which reward center stimulation could kill.

**Wireheading to death**. Experiments on rats showed that a rat (Olds and Milner 1954) will constantly press a lever to get the next stimulating impulse. It is often reported that rats will die because of starvation but this was not actually measured in Olds and Milner’s experiments, as the rats were artificially killed for examination. The only experiment where rats actually starved themselves is described by Routtenberg (Routtenberg 1964), and the experimenters only measured weight loss, as they needed the rats alive at the end of the experiment for anatomic analyses (Routtenberg and Lindy 1965).

In an experiment by Johanson et al., half of monkeys which intravenously self-administrated different drugs died from different forms of overdosing in 3-5 days, but others survived 30 days to the end of the experiment and continued to eat to some extent: “For Animal 4029, drug intake varied over the 29-day access period, but did not seem to increase with continued access. Although food intake was initially suppressed, it returned to predrug levels by Day 13. Food consumption continued to fluctuate from 50 to 200 g over the next 16 days; however, food and drug intake were not systematically related” (Johanson et al. 1976). Another animal likely died of starvation on a stimulant drug: “However, not only was food intake initially suppressed when access to drug began, but consumption never returned to more than 20 percent of predrug levels” (Johanson et al. 1976). “Rhesus monkeys given unlimited access to a variety of psychomotor stimulant drugs self-administer them in amounts sufficient to result in death after less than 3 weeks of availability. Most of the animals who had access to cocaine, d-amphetamine, l-amphetamine and d-methamphetamine survived less than 5 days in the 23 hr/day regimen and never more than 15 days except for one animal on 1-amphetamine. Three of the 5 animals will access to diethylpropion, however, survived for the duration of the 28-30 day experiment, despite, in 2 of the animals, high levels of daily intake up to 196 mg/kg” (Johanson et al. 1976). In a review of these and similar studies, Wise concluded: “Rats and monkeys will work similarly compulsively for intravenous stimulants; if given unlimited access, they will self-administer intravenous injections of these drugs to the point of severe weight loss and death” (Wise 2002).

In other words, the combination of starvation and overdose was deadly for the majority of animals, but some of them clearly died of overdose, and others from starvation.

**Overdoses** are the main cause of death from opiates, largely because this class of drug affects the breathing control centers within the brain (Maron 2018). Suicides are also very high in heroin users (Darke and Ross 2002). The illegal market is constantly changing the concentration of the active ingredient of the drug, increasing chance of overdose. Some drugs are substituted for or mixed with others with different effects.

One can imagine that pure wireheading will pose less of a risk of overdose, as it would not affect the breathing control center. But almost any addiction has some deadly mode (overeating – heart disease, sex – sexually transmitted infections (STIs), and so on), the reason for which is that addiction is a self-accelerating process. There is a slippery slope to seeking greater and greater pleasures, and at some point, such a self-accelerating process reaches a point where it affects some life-supporting system.

**Lower life expectancy**. Many addictive drugs and activities lower life expectancy via different mechanisms. The route of administration may be damaging in the case of injections. The long-term side-effects of toxicity may affect health. Unhealthy life habits and self-neglect also shorten lifespans.

**Damage to reward function**. Intensive stimulation of the reward function lowers a person’s ability to be stimulated by anything else (Garfield et al. 2014), which creates anhedonia accompanied by depression, social withdrawal, non-adaptive behavior, less self-care and shorter life (Blanchard et al. 1998) Even trying something very rewarding once could create a new psychological “beacon” in the reward function that a person would consciously or subconsciously pursue. Memory of an intense pleasure above some threshold is, in some sense, similar to post-traumatic stress disorder, as it creates compulsive memories of the event. The human reward function is able to self-adapt by providing different rewards for different activities that are needed for the organism at a specific moment. For example, if glucose is low, it will provide a higher reward for eating, and lower the reward after getting signals from a full stomach. But artificial wireheading could damage this self-adapting capability of the human reward center, which will damage basic physiological functions as well as complex social behavior.

**Socially dangerous behavior**. People under the effect of drugs may act dangerously and irrationally. Most typically this behavior appears under the influence of alcohol, but it can also occur with amphetamines and other drugs (e.g. bath salts, PCP). Cravings during withdrawal may cause addicts to commit serious crimes for small amounts of money. A drug addict becomes something similar to what Bostrom called “paperclip maximizer” (an AI which maximizes a completely useless utility function (Bostrom 2014)). Crimes are also committed by drug-dealing gangs.

**Painful withdrawal and risking toxic cures** like the lethal combination of some illicit drugs and antidepressants (Dobry et al. 2013) is another failure mode.

**Lower fertility and damage to children**. Some drugs damage interest in reproduction, reproduction capabilities, and health of the children of users (e.g. fetal alcohol syndrome) (Wilson Jones and Thomas Bass 2003).

**Intelligence decline**. Not only the human value system is damaged by drugs, but also rationality and other cognitive abilities (Goldstein et al. 2009).

**Fueling of malicious biohackers**. Home-made drug production would fuel DIY biohackers and create instruments for localized biotech production. Moreover, it would fuel widespread access to genetic manipulation technology, that is, the capability to generate custom DNA at home after downloading its code from the internet. The DNA is installed in some standard carrier, like *E. coli* or yeast; this is especially likely if the universal bio-synthesizers invented by Craig Venter become popular (Boles et al. 2017).

In this scenario, a widely available biotechnology could give access to all possible illegal drugs to each person, in the same way in which one could watch every movie that has ever existed via pirate torrents (Morris 2008). Surely, many “biohackers” would be interested in having such labs, but the same labs could be used to synthesize dangerous toxins or even viruses, if, for example, they are remotely hacked (Turchin et al. 2017). In other words, proliferation of the illegal biosynthesis labs would likely be fueled by demand for different recreational drugs, but a large network of such labs increases the chance that they will be used for toxin or bioweapons production.

A new risk proposed here is that the loss of control of euphoric drug-producing bacteria could start a “happiness pandemic”, for example, by infecting gut microbiomes. The problem with its containment would be that its carriers will likely be happy to share their disease. The result would be something like a zombie-pandemic. Some viral vectors already promote their own dissemination in mammals (e.g. rabies and toxoplasmosis) by inducing behavioral changes in their hosts (Flegr 2007).

**Environmental pollution by genetically modified organism (GMO) bacteria and plants.** Drug-producing homemade GMO bacteria/fungi/plants could leak drugs into the environment and result in accidental overdoses in animals or people. This is unlikely.

**Health risks of “dirty” wireheading**. In search of a quick depression fix, permanent capability amplification or as a one-time payment for permanent access to a high, some people might pay for illegal wireheading. The procedure could be relatively simple and inexpensive if done without many medical precautions. This “dirty” wireheading could result in eventual long-term damage to neuronal pathways. Electrodes tend to degrade over time; another risk is that increased electric stimulation could eventually kill neurons or cause other types of neurodegeneration.

Even the best medical procedure of deep brain stimulation involves risks: a 3 per cent chance of brain hemorrhage (an insult which itself could cause permanent brain damage) and a 3–5 per cent risk of infection which would require secondary surgery to remove the device (Tolleson et al. 2014). Dirty wireheading done at non-medical facilities would probably have even higher risks of complications. Taylor mentioned that too many regulations for safe medical wireheading would only fuel a black market (Taylor 2012). She suggests that the highest available level of stimulation should be legally limited and that some measures like temporary turn-off should be employed to prevent a slippery slope to higher levels of simulation. People who are already addicted to an opioid high may prefer a one-time payment and the risk of wireheading to the constant risks of drug overdose.

From this list, it could be concluded that while the idea that “wireheading will cause inevitable death by starvation” is not true, the multiplicity of killing mechanisms of wireheading is large enough to possibly become a global threat.

## 4.2 Catastrophic wireheading criteria

As we show above, wireheading could be dangerous, but alone it is not enough to cause a global catastrophic risk: for instance, most people do not try hard drugs despite knowing of their existence and possessing the ability to obtain them. Future advances in neurostimulation could also presumably create “good wireheading” which may be able to stop suffering without creating health risks, and which will be discussed in the next section.

A “super-drug” needs an ability to act globally, for the duration of a human life-time and affect large part of the human population to become a global catastrophic risk:

* **Based on knowledge**, not be a material object, or its material part should be easily accessible. For example, games involving sexual asphyxia are based on pure knowledge, and this information can be disseminated through any informational channel. More importantly, they can be disseminated through collective practice, like the choking game (Linkletter et al. 2010). If future wireheading is just an adjustment of existing commercial neuroimplants, it would also be purely knowledge-based. Databases like erowid.com serve as marketplaces for the exchange of new information (Rothstein 2015).
* **Simple to use**. It is known that availability strongly affects the rate of drug self-administration. It is difficult not to eat if you have fridge full of addictive food.
* **Slippery slope**. An atypical antidepressant, tianeptine, was eventually banned in Russia as an increase in dose creates gradually increasing euphoria (Springer and Cubała 2018). Such a slippery slope creates a death spiral seeking ever-higher levels of stimulation.
* **Promotes its sharing**. Such a drug has an internal mechanism not only of reward stimulation, but promoting its dissemination to other people. For example, an addict is motivated to give a first dose of his/her drug to his/her friends, so they will also become addicts and s/he will sell them the drug, much like a multilevel marketing scheme.
* **Non-obvious long-term effects**. Quick death following wireheading will be obviously bad and other people would choose to avoid wireheading due to its obvious risk. Very slow death would also allow time for social adaptation and coping measures to develop. Something which becomes widespread before its potential for danger is recognized, maybe because of the slippery slope, could be a real danger.
* **Not bad at first glance**. This is a risk which does not trigger an immediate warning or is not obviously bad. Nearly everybody knows that starting heroin is the road to hell, with a high probability of death in a few years. But it is not so obvious if one just takes the first pill of a prescribed opiate pain killer, or, say, a new brain stimulant. Some stimulants, like amphetamines, initially increase social adaptation (Pedersen et al. 2015).
* **Super-drug overwhelms society’s ability to cope** with and control it. If the danger of some form of wireheading is recognized, the same drug police (e.g. Drug Enforcement Administration) as well as advertising campaigns will be used to stop it. However, if the ease to obtain and incentive to abuse the drug were high, it could overwhelm coping ability, like alcohol abuse which continued during Prohibition in the U.S.
* **Connected with an ideology** like short-term hedonism, nihilism, or some form of hedonic utilitarianism. Humans could create supporting ideologies to find rationale for their addictions, like some people praise marijuana as a universal cure and others drink because they think that this will clean their blood vessels (Grimes 2018). There are philosophical publications about the possible benefits of wireheading done correctly, e.g. (Taylor (2012), (Muflax 2011), the site *wireheading.com*. In contrast, there are almost no publications about the catastrophic risk of wireheading in English (though some exist in Russian (Strugatsky and Strugatsky 1976, hitthelimit 2008, Argonov 2015)).

The main difference between catastrophic wireheading technology and other global catastrophic risks is that in the case of wireheading, humans would generally not be averse to succumbing to the danger and could even actively participate in its promotion.

The second difference is that in the case of other risks, different small constituents could have cumulative negative effect (like many viral pandemic simultaneously could increase probability of human extinction (Turchin et al. 2018)). For wireheading, a variety of possible artificial pleasures creates some form of protection, because searching for different pleasures is similar to a typical human life; most people would prefer to live a diverse life and would prefer not to succumb to simple button pressing.

## 4.3 Positive world of eternal pleasures

It is possible that civilization could create a world of almost immortal people with indefinitely many interesting activities and pleasures, as described in Yudkowsky’s “general theory of fun” (Yudkowsky 2009). However, if these pleasures are slightly too strong, the result could be a global wireheading dystopia.

The risks of bad wireheading should not stop humanity from attempts to augment the human mind to escape unbearable suffering, depression, and aggression. As Taylor mentioned, some form of mild hypomania may be beneficial (Taylor 2012). However, But such wireheading should be self-administrated under control of some paternalistic entity to lower risks.

Good wireheading is something like brain stimulation + nootropics; one example is coffee, which not only improves quality of life, but based on recent research, extends life expectancy (Gunter et al. 2017). Good wireheading will probably include the intellectual pleasure of complex problem solving, many different pleasant activities, an increase in social adaptation and compassion, domination of long-term planning, and improvement of brain health. Positive wireheading should be good enough but not too good. It is great to be in a good working mood, where you are in the flow and every task is easy, but if one feels “too good”, one will be able only to perform “trainspotting”, that is mindless staring at objects.

Good wireheading should have the right slope of the pleasure gradient, that is, subjective increase of pleasure with increase of a dose. If the pleasure gradient is too steeply sloped, one would want even more reward despite already experiencing high reward. If the pleasure gradient has something like a local maximum, then something which feels “too good” would trigger a fear of losing reaction or other back reaction, and an individual would self-calibrate to the local optimum.

In the Strugtaskys’ novel (Strugatsky and Strugatsky 1976), “sleg” is not just wireheading in the form of a perfect pleasure: it is something like a controlled lucid dream, so pleasant and entertaining that people may kill others who interrupt them. This outcome was also envisioned in Wenders’ movie “Until the end of the world,” in which the main character becomes addicted to recordings of her own dreams (Wenders 1991). A “Sleg”-like addiction to virtual worlds may be similar to a computer game, but these worlds would be unlike multiplayer games that create a new form of social activity. Multiplayer games can even help people escape dangerous chemical addictions (Clune 2015).

# 5 Wireheading in the context of other catastrophic risks and civilizational problems

“... Zhilin,” said Rimayer, “history is a story of people. Every person wants to live life with good reason, and gives you such a slip ... Yes, I know, you think that you live without a trace for good reason, but confess, you have never lived so brightly and hotly as you do in the bathtub today. Are you a little ashamed to remember, would you dare to tell others about this life? Do not need. They have their own lives, you have your own ...”

“... Rimayer,” Zhinlin said, “all this is true. But the past! Space, schools, the fight against the fascists, with gangsters - what, all this in vain? Forty years I have lived in vain? And the others? Too in vain? ..

“... Zhilin,” said Rimayer, “in history, nothing is wasted. Some fought and did not live to see the sleg. And you fought and lived ...”

*Strugarskys’ “The final circle of the paradise”*

## 5.1 Relation between drug legalization, technological unemployment and basic income

The current trend in many Western countries is the legalization of marijuana, and the next probable candidates are psychedelics like psilocybin and ketamine, which are promoted as effective treatments for depression (Newport et al. 2015). Future civilization will likely have many unemployed people supported by some form of basic income or welfare. Such people will not have as much business to occupy their time and will likely turn to more inexpensive reward stimulation: now they have hobbies/sports, watch TV, use the Internet, play computer games, and drink, but later they may consume newly legalized drugs.

## 5.2 Could widespread wireheading be an explanation of the Fermi paradox?

Wireheading, together with recognition of ontological nihilism (Hitthelimit 2008), that is, the idea that no goal is intrinsically better than any other goal, could presumably stop the further development of civilization. Why fly to the stars, if you could live in simulations or even just live in infinite bliss via reward function hacking?

However, the power of natural selection would probably counteract such an outcome. While some people will likely succumb to wireheading, others may be especially reluctant to be reward-stimulated. We can see that even now, most people are just not interested in starting to take opiates, even if they know about them and could obtain them. On the other hand, availability strongly influences consumption, and if pressing a button could be equal to receiving a reward burst, nearly everyone will sometimes have an urge to reward themselves.

Some minds might develop a control system which is less reward-based and more rule-based—"do what you ought”—and long-term natural selection will overcome any new ways of reward stimulation. The main question is, will humanity have enough time for such adaptation to take place?

We define a “reward-stimulating technology overhang” as …. In this case, even social collapse may not stop the usage of the high-tech brain stimulation methods, for example, if some GMO-drug producing plants survive.

## 5.3 Possible protective measures

Traditional legal restrictions could be used for punishment and to create a stigma associated with new ways of reward-hacking. A radical new way to eliminate risk proposed here is a human brain upgrade which prevents unlimited wireheading, so humans could explore the world of possible pleasures without the risk of addiction.

There is a suggestion that humans are more robust to direct wireheading than animals (https://www.wireheading.com/), as they are able to stop self-administrating cocaine for eating and sleep, while other primates have only limited control (and rats have no control). In other words, humans do have some “rule-based” willpower in the neocortex that is able to—to some extent—override reward system signals.

Human brains are diverse, and some people may have a weaker tendency toward addiction, which, among other factors, depends on variation in the dopamine receptors. Many different types of possible addictions could create some form of complex ecology, not much different from current society, where different attractions are fighting for human attention.

Widespread wireheading is unlikely to cause human extinction as it is difficult to imagine that everybody will do the same thing at the same time, and that one activity would affect everybody at the same time. However, wireheading could contribute to the collapse of a society and be a part of the more complex process which will result in civilizational collapse, which itself is a global catastrophic risk (Turchin and Denkenberger 2018b) and could later turn into an extinction event (Denkenberger and Pearce 2017).

Future technological growth will create many new ways to stimulate human reward centers, and if these methods arise too quickly, society may not have time to cope with the changes and create adequate social rituals and norms to canalize such activities into prosperity as has been done in the past. For example, some societies have long traditions of wine consumption, like France, where it may even contribute to an increase in life expectancy, while other societies, like Northern native peoples, could be almost wiped out by vodka.

# Conclusion

The speed of development of human reward-hacking technologies depends on the speed of progress in AI neuroscience, neuroimplants, and synthetic biology. Given the current high speed of progress in these fields, an explosion of brain-stimulating technologies could occur in the next 10–30 years. Humanity will be more and more effective in its reward stimulation, which will be cheaper and often depend only on knowledge. This combination of factors could overcome society’s coping mechanisms and result in its decline. More research should be done on preventing this decline.

**Disclaimer**

This article represents views of the authors and does not necessarily represent the views of the Global Catastrophic Risk Institute or the Alliance to Feed the Earth in Disasters. No external sources of funding were used for this work.

# Literature

1. ABC. 2017. Back Pain Device Produces Orgasms - ABC News. *ABCnews*.
2. Adewole, D. O., L. A. Struzyna, J. P. Harris, A. D. Nemes, J. C. Burrell, D. Petrov, R. H. Kraft, H. I. Chen, M. D. Serruya, J. A. Wolf, and D. K. Cullen. 2018. Optically-Controlled “Living Electrodes” with Long-Projecting Axon Tracts for a Synaptic Brain-Machine Interface. *bioRxiv*:333526.
3. Amihai, I., and M. Kozhevnikov. 2015. The influence of Buddhist meditation traditions on the autonomic system and attention. *BioMed research international* 2015.
4. Andreassen, C. S. 2015. Online social network site addiction: A comprehensive review. *Current Addiction Reports* 2(2):175–184.
5. Argonov, V. 2015. Rethinking progress: beyond the horizon.
6. Baeshen, N. A., M. N. Baeshen, A. Sheikh, R. S. Bora, M. M. M. Ahmed, H. A. I. Ramadan, K. S. Saini, and E. M. Redwan. 2014. Cell factories for insulin production. *Microbial Cell Factories* 13.
7. Boles, K. S., K. Kannan, J. Gill, M. Felderman, H. Gouvis, B. Hubby, K. I. Kamrud, J. C. Venter, and D. G. Gibson. 2017. Digital-to-biological converter for on-demand production of biologics. *Nature Biotechnology, 35, 672–675 (2017)*.
8. Bostrom, N. 2014. *Superintelligence*. Oxford University Press, Oxford.
9. Browne, R. 2011, April 9. A fine line to an early death. *The Sydney Morning Herald*.
10. Carlezon Jr, W. A., and E. H. Chartoff. 2007. Intracranial self-stimulation (ICSS) in rodents to study the neurobiology of motivation. *Nature protocols* 2(11):2987.
11. CDC. 2018, May 9. Smoking and Tobacco Use; Fact Sheet; Tobacco-Related Mortality. http://www.cdc.gov/tobacco/data\_statistics/fact\_sheets/health\_effects/tobacco\_related\_mortality/.
12. Cox, J. 2013, August 14. The Woman Who Drilled a Hole in Her Head to Open Up Her Mind. *Vice*.
13. Darke, S., and J. Ross. 2002. Suicide among heroin users: rates, risk factors and methods. *Addiction* 97(11):1383–1394.
14. Delight Pro. 2018. Store - Mind Alive. https://mindalive.com/index.cfm/store/product.cf?product=6.
15. Disner, S. G., C. G. Beevers, and F. Gonzalez-Lima. 2016. Transcranial Laser Stimulation as Neuroenhancement for Attention Bias Modification in Adults with Elevated Depression Symptoms. *Brain Stimulation* 9(5):780–787.
16. Dobry, Y., T. Rice, and L. Sher. 2013. Ecstasy use and serotonin syndrome: a neglected danger to adolescents and young adults prescribed selective serotonin reuptake inhibitors. *International Journal of Adolescent Medicine and Health* 25(3):193–199.
17. Docimo, T., M. Reichelt, B. Schneider, M. Kai, G. Kunert, J. Gershenzon, and J. C. D’Auria. 2012. The first step in the biosynthesis of cocaine in Erythroxylum coca: the characterization of arginine and ornithine decarboxylases. *Plant Molecular Biology* 78(6):599–615.
18. Drexler. 1986. *E.: Engines of Creation*. Anchor Press.
19. Fagundes, V. de C., C. R. M. Rieder, A. N. da Cruz, B. C. Beber, and M. W. Portuguez. 2016. Deep Brain Stimulation Frequency of the Subthalamic Nucleus Affects Phonemic and Action Fluency in Parkinson’s Disease. *Parkinson’s Disease* 2016.
20. Flegr, J. 2007. Effects of Toxoplasma on Human Behavior. *Schizophrenia Bulletin* 33(3):757–760.
21. Frank, R. G., and H. A. Pollack. 2017. Addressing the Fentanyl Threat to Public Health. *New England Journal of Medicine* 376(7):605–607.
22. Franson, J. 2013. Transcranial direct current stimulation tDCS technology may improve cognitive ability | National Post. *National Post*.
23. Galanie, S., K. Thodey, I. J. Trenchard, M. Filsinger Interrante, and C. D. Smolke. 2015. Complete biosynthesis of opioids in yeast. *Science (New York, N.Y.)* 349(6252):1095–1100.
24. Gilliam, T. 1998. Fear and Loathing in Las Vegas. Adventure, Comedy, Drama.
25. Global Market Insights. 2017. Neurostimulation Devices Market worth over $13 billion by 2023: Global Market Insights, Inc. https://globenewswire.com/news-release/2017/07/12/1042929/0/en/Neurostimulation-Devices-Market-worth-over-13-billion-by-2023-Global-Market-Insights-Inc.html.
26. Gonzalez, R. 2018, April 6. A Brain-Boosting Prosthesis Moves From Rats to Humans. *Wired*.
27. Goodman, B. 2011. Deep Brain Stimulation May Offer Lasting Benefits for Parkinson’s Disease. *WebMed*.
28. Groppa, S., A. Oliviero, A. Eisen, A. Quartarone, L. G. Cohen, V. Mall, A. Kaelin-Lang, T. Mima, S. Rossi, and G. W. Thickbroom. 2012. A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clinical Neurophysiology* 123(5):858–882.
29. Gunter, M. J., N. Murphy, A. J. Cross, L. Dossus, L. Dartois, G. Fagherazzi, R. Kaaks, T. Kühn, H. Boeing, K. Aleksandrova, A. Tjønneland, A. Olsen, K. Overvad, S. C. Larsen, M. L. Redondo Cornejo, A. Agudo, M. J. Sánchez Pérez, J. M. Altzibar, C. Navarro, E. Ardanaz, K.-T. Khaw, A. Butterworth, K. E. Bradbury, A. Trichopoulou, P. Lagiou, D. Trichopoulos, D. Palli, S. Grioni, P. Vineis, S. Panico, R. Tumino, B. Bueno-de-Mesquita, P. Siersema, M. Leenders, J. W. J. Beulens, C. U. Uiterwaal, P. Wallström, L. M. Nilsson, R. Landberg, E. Weiderpass, G. Skeie, T. Braaten, P. Brennan, I. Licaj, D. C. Muller, R. Sinha, N. Wareham, and E. Riboli. 2017. Coffee Drinking and Mortality in 10 European Countries: A Multinational Cohort Study. *Annals of Internal Medicine* 167(4):236.
30. Hanson, R. 2008. Catastrophe, social collapse, and human extinction. Page 554 *in* N. Bostrom and M. M. Cirkovic, editors. *Global catastrophic risks (p*. Oxford University Press, Oxford.
31. Hanson, R. 2016. *The Age of Em: Work, Love, and Life when Robots Rule the Earth*. Oxford University Press.
32. hitthelimit. 2008. Психогенная сингулярность. LJ.
33. Johanson, C. E., R. L. Balster, and K. Bonese. 1976. Self-administration of psychomotor stimulant drugs: The effects of unlimited access. *Pharmacology Biochemistry and Behavior* 4(1):45–51.
34. Khalturina, D. A., and A. V. Korotaev. 2008. Alcohol and narcotics as factors of the demographic crisis. *Sociological Research* 47(3):18–31.
35. Lesswrongwiki. 2018. Orgasmium - Lesswrongwiki.
36. Linkletter, M., K. Gordon, and J. Dooley. 2010. The choking game and YouTube: a dangerous combination. *Clinical pediatrics* 49(3):274–279.
37. Maron, D. F. 2018. How Opioids Kill. *Scientific American*.
38. Merrer, J. L., J. A. J. BECKER, K. BEFORT, and B. L. KIEFFER. 2009. Reward Processing by the Opioid System in the Brain. *Physiological reviews* 89(4):1379–1412.
39. Miller, J. D. 2018. A Rationally Addicted Artificial Superintelligence.
40. Muflax. 2011. Why No Wireheading?
41. National Center for Health Statistics. 2018, March 5. FastStats. https://www.cdc.gov/nchs/fastats/deaths.htm.
42. National institute on drug abuse. 2018. Overdose Death Rates | National Institute on Drug Abuse (NIDA). https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates.
43. Nazaretian, A. 2004. *Civilizationa crisises in the context iof universal history. (In Russian)*. Mir.
44. Newport, D. J., L. L. Carpenter, W. M. McDonald, J. B. Potash, M. Tohen, C. B. Nemeroff, A. C. of R. T. F. on N. Biomarkers, and Treatments. 2015. Ketamine and other NMDA antagonists: early clinical trials and possible mechanisms in depression. *American Journal of Psychiatry* 172(10):950–966.
45. Neyaskin, G. 2018, January 27. Reality Check: Fall in Russian alcohol consumption. *BBC News*.
46. Oberst, U., E. Wegmann, B. Stodt, M. Brand, and A. Chamarro. 2017. Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *Journal of adolescence* 55:51–60.
47. Olds, J., and P. Milner. 1954. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of comparative and physiological psychology* 47(6):419.
48. Oster, G. 1973. Auditory beats in the brain. *Scientific American* 229(4):94–103.
49. Pedersen, W., S. Sandberg, and H. Copes. 2015. High speed: Amphetamine use in the context of conventional culture. *Deviant Behavior* 36(2):146–165.
50. Pierre, L. S.-, and M. A. Persinger. 2006. Experimental facilitation of the sensed presence is predicted by the specific patterns of the applied magnetic fields, not by suggestibility: re-analyses of 19 experiments. *International Journal of Neuroscience* 116(19):1079–1096.
51. Polianskaia, A. 2018, July 7. Jackson Coe found dead at bottom of building in New York. *The Independent*.
52. Portenoy, R. K., J. O. Jarden, J. J. Sidtis, R. B. Lipton, K. M. Foley, and D. A. Rottenberg. 1986. Compulsive thalamic self-stimulation: a case with metabolic, electrophysiologic and behavioral correlates. *Pain* 27(3):277–290.
53. Price, L. 2018. Neural implants and the race to merge the human brain with Artificial Intelligence. *Thought Leadership In Digital Health | London | UK*.
54. Ramos, L., J. E. Hunter, and J. H. Wolfe. 2017. Viral Vector Gene Delivery to the Brain for Treating Neurogenetic Diseases. Pages 89–125 *Drug and Gene Delivery to the Central Nervous System for Neuroprotection*. Springer.
55. Reuters. 2007, September 17. Man in China dies after three-day Internet session. *Reuters*.
56. Rothstein, A. 2015, November 25. How the Most Extreme Trips on Erowid Transformed Modern Drug Culture. *Motherboard*.
57. Routtenberg, A. 1964. SELF-STARVATION CAUSED BY FEEDING CENTER SELF-STIMULATION. Pages 502–502 *American Psychologist*. AMER PSYCHOLOGICAL ASSOC 750 FIRST ST NE, WASHINGTON, DC 20002-4242.
58. Routtenberg, A., and J. Lindy. 1965. Effects of the availability of rewarding septal and hypothalamic stimulation on bar pressing for food under conditions of deprivation. *Journal of comparative and physiological psychology* 60(2):158.
59. Schlaepfer, T. E., B. H. Bewernick, S. Kayser, B. Mädler, and V. A. Coenen. 2013. Rapid Effects of Deep Brain Stimulation for Treatment-Resistant Major Depression. *Biological Psychiatry* 73(12):1204–1212.
60. Simler, K., and R. Hanson. 2017. *The Elephant in the Brain: Hidden Motives in Everyday Life*. Oxford University Press.
61. Sparrow, G., R. Hurd, R. Carlson, and A. Molina. 2018. Exploring the effects of galantamine paired with meditation and dream reliving on recalled dreams: Toward an integrated protocol for lucid dream induction and nightmare resolution. *Consciousness and Cognition* 63:74–88.
62. Springer, J., and W. J. Cubała. 2018. Tianeptine Abuse and Dependence in Psychiatric Patients: A Review of 18 Case Reports in the Literature. *Journal of Psychoactive Drugs* 50(3):275–280.
63. Strugatsky, A., and B. Strugatsky. 1976. The Final Circle of Paradise. *Trans. Leonid Renen. New York, DAW*.
64. Suthar, A. B., and T. Bärnighausen. 2017. Antiretroviral therapy and population mortality: Leveraging routine national data to advance policy. *PLoS Medicine* 14(12).
65. Taylor, J. 2012. Wireheading: Towards a Consumer Market?
66. Templeton, G. 2017. Elon Musk’s NeuraLink Is Not a Neural Lace Company. https://www.inverse.com/article/30600-elon-musk-neuralink-neural-lace-neural-dust-electrode.
67. Toblin, R. L., L. J. Paulozzi, J. Gilchrist, and P. J. Russell. 2008. Unintentional strangulation deaths from the" Choking Game" among youths aged 6-19 years–United States, 1995—2007. *Journal of safety research* 39(4):445–448.
68. Tolleson, C., J. Stroh, J. Ehrenfeld, J. Neimat, P. Konrad, and F. Phibbs. 2014. The factors involved in deep brain stimulation infection: a large case series. *Stereotactic and Functional Neurosurgery* 92(4):227–233.
69. Tomasik, B. 2015. How Likely is Wireheading?
70. Turchin, A., and D. Denkenberger. 2018. Classification of Global Catastrophic Risks Connected with Artificial intelligence. Under review in AI&Society.
71. Turchin, A., B. Green, and D. Denkenberger. 2017. Multiple Simultaneous Pandemics as Most Dangerous Global Catastrophic Risk Connected with Bioweapons and Synthetic Biology. *Under review in Health Security*.
72. Van Besouw, R. 2013. Implanting Awareness |. https://www.soundonsound.com/sound-advice/implanting-awareness.
73. Wenders, W. 1991. Until the End of the World. Action, Drama, Sci-Fi.
74. Wise, R. A. 2002. Brain Reward Circuitry: Insights from Unsensed Incentives. *Neuron* 36(2):229–240.
75. Wood, A. M., S. Kaptoge, A. S. Butterworth, P. Willeit, S. Warnakula, T. Bolton, E. Paige, D. S. Paul, M. Sweeting, and S. Burgess. 2018. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *The Lancet* 391(10129):1513–1523.
76. Yager, L. M., A. F. Garcia, A. M. Wunsch, and S. M. Ferguson. 2015. The ins and outs of the striatum: Role in drug addiction. *Neuroscience* 301:529–541.
77. Yampolskiy, R. 2014. Utility Function Security in Artificially Intelligent Agents. *Journal of Experimental and Theoretical Artificial Intelligence (JETAI)*:1–17.
78. Yudkowsky, E. 2009. The Fun Theory Sequence - Less Wrong.