systems

MDPI

Concept Paper

Ethical Regulators and Super-Ethical Systems

Mick Ashby [1,2,3]

[1] W. Ross Ashby Digital Archive, London, UK; ethics@ashby.de
[2] Cybernetics Society, London W5 2NQ, UK
[3] American Society for Cybernetics, Washington, DC 20052, USA

Received: 2 November 2020; Accepted: 3 December 2020; Published: 9 December 2020

Abstract: This paper combines the good regulator theorem with the law of requisite variety and seven other requisites that are necessary and sufficient for a cybernetic regulator to be effective and ethical. The ethical regulator theorem provides a basis for systematically evaluating and improving the adequacy of existing or proposed designs for systems that make decisions that can have ethical consequences; regardless of whether the regulators are humans, machines, cyberanthropic hybrids, organizations, or government institutions. The theorem is used to define an ethical design process that has potentially far-reaching implications for society. A six-level framework is proposed for classifying cybernetic and superintelligent systems, which highlights the existence of a possibility-space bifurcation in our future time-line. The implementation of "super-ethical" systems is identified as an urgent imperative for humanity to avoid the danger that superintelligent machines might lead to a technological dystopia. It is proposed to define third-order cybernetics as the cybernetics of ethical systems. Concrete actions, a grand challenge, and a vision of a super-ethical society are proposed to help steer the future of the human race and our wonderful planet towards a realistically achievable minimum viable cyberanthropic utopia.

Keywords: ethical AI; superintelligence; empathy; sapience; utopia; third-order cybernetics

1. Introduction

The goal of this research is to develop a theoretical basis and a practical systematic process for designing systems that behave ethically; even under non-ideal or hostile conditions.

The human race has become very good at designing systems that are effective, but we are very bad at designing systems that are reliably ethical. The majority of our social and computer-based systems are ethically fragile, lacking resilience under non-ideal conditions, and are generally vulnerable to abuse and manipulation. But we are now on the cusp of a technological wave that will thrust autonomous vehicles, robots, and other artificial intelligence (AI) systems into our daily lives, for good or bad; there will be no stopping them. And despite widespread recognition of the potential risks of creating superintelligence [1] and the need to make AI and social systems ethical, systems theory, cybernetics, and AI have no adequate systematic processes for creating systems that behave ethically. Instead, we have to rely on the ad hoc skills of an ethically motivated designer to somehow specify a system that is hopefully ethical, despite the constant pressure from corporate executives to do things cheaper and faster. This is not a satisfactory solution to a problem that so urgently needs to be solved. In the context of cybernetics, this could be referred to as "The Ethics Problem".

Many people think that all technologies can be used for good or evil, but this is not true. If we consider a system like that of public health inspections of restaurants, where an inspector performs a well-structured system of evaluations in defined dimensions, such as kitchen hygiene, food storage, waste management, and signs of vermin, to identify any inadequacies and specify necessary improvements to achieve certification of hygienic adequacy; such a system can only help to make restaurants more hygienic.

Might it be possible to adapt this certification model from the public health domain to create a system that can be used to certify whether a given system is ethically adequate or inadequate? And might such a system be a solution to "The Ethics Problem"?

In this paper, the terms "ethics" and "ethical" are used in a concrete applied sense of the acceptable behaviour in a society or nation. Treating ethics as an absolute standard that can be applied across all cultures is rejected, subscribing rather to Manel Pretel-Wilson's assertion that "there is no supreme value but a plurality of values." [2]. For some systems, the term "ethical" might include aspects such as hygienic, safe, fair, honest, law-abiding, or environmentally friendly.

All societies regulate the behaviour of their members by defining what behaviour is acceptable or unacceptable. It is primarily through the rule of law that a society can be made safe, civilized, and "ethical" as defined by the norms of that society. And the only way that society or an individual can know or prove that something non-trivially unethical has occurred is because some kind of rule has been violated. So being pragmatic, if it is unethical to break laws, regulations, and rules, then it is laws, regulations, and rules that define the ethics for each legislative jurisdiction, which is why we bother to constantly try to improve them. Not all rules are defined formally in writing, some are unwritten conventions, yet in every culture, it is unacceptable to break such laws, regulations, rules, or customs.

However, the act of deciding what is ethical behaviour is very different from the act of behaving ethically by obeying a society's laws and rules. It is the lawmakers who face genuine ethical dilemmas when making decisions about which behaviours are to be legislated as acceptable in society or forbidden. But a law-abiding citizen (or machine) needs only to obey the appropriate laws and rules in order to behave safely and ethically in all anticipated situations, with an acceptably small risk that something unethical might result despite following the current laws and rules.

Just as a law-abiding citizen does not need to be involved in the ethical decisions that are required when making laws, this paper does not address the issue of how a society decides what is made legal or illegal. We must accept that a law-abiding machine, like many law-abiding citizens, might blindly obey unjust or evil laws, such as those made in Nazi Germany, but the problems of the existence of evil dictatorships and unethical laws must be addressed at the political level. That is something that each society must resolve for itself and is outside the scope of this paper, which is concerned rather with how to create effective systems that are certifiably and reliably law-abiding; even under non-ideal or hostile conditions.

None of us are ever likely to have to decide whether to switch a runaway train to a different track to reduce the number of fatalities, but if the lawmakers of a country decide, for example, that in such situations, minimizing fatalities is the ethical and legal obligation, then it becomes trivial to encode it in a law, regulation, or rule so that it can be understood and obeyed by humans and machines. By doing so, what was an ethical dilemma is reduced to a simple rule. This line of reasoning implies that it is sufficient to disambiguate our laws and make robots, artificial intelligence, and autonomous vehicles rigorously law abiding. It is suggested that there is absolutely no need to make such autonomous systems capable of resolving genuine ethical dilemmas, which is the job of society's lawmakers and regulatory organizations to anticipate, resolve, and codify in advance.

Methodology

The starting point for this research was trying to find a complete set of answers to the question "What properties must a system have for it to behave ethically?".

The existing cybernetics literature provided the first two properties. Roger Conant and Ross Ashby's Good Regulator Theorem [3] proved that every good regulator of a system must be a model of that system, but it does not specify *how* to create a good regulator. And Ross Ashby's Law of Requisite Variety [4] dictates the range of responses that an effective regulator must be capable of. However, having an internal model and a sufficient range of responses is insufficient to ensure effective regulation, let alone ethical regulation. An ethical system must have more than just these two properties.

Recent approaches to making AI ethical, such as IBM's "Everyday Ethics for Artificial Intelligence: A practical guide for designers and developers" [5] and the European Commission's "High-Level Expert Group on Artificial Intelligence: Draft Ethics Guidelines for Trustworthy AI" [6] provided "ethically white-washed" [7] lists of requirements, without offering anything that could be applied systematically and repeatably to actually design an ethical AI.

Heinz von Foerster proposed an ethical imperative: "Act always so as to increase the number of choices" [8]. Although this principle is valuable in the context of psychological therapy, it specifies no end condition, i.e., when to stop adding more choices. If one were to apply it when deciding how many different types of propulsion systems to build into a manned-spacecraft to adjust its motion and orientation, it would lead to unnecessary choices, unnecessary weight, unnecessary costs, increase the number of points of possible failure, and therefore increase the risk of catastrophic failure and loss of life. This counter example proves that maximizing choice can be the wrong (unethical) thing to do. And by definition, implementing more choices than is necessary to achieve the goal of a system is unnecessary and violates the principle of Occam's razor. So we must reject von Foerster's Ethical Imperative as being flawed.

In 1990, von Foerster gave a lecture titled "Ethics and Second-Order Cybernetics" to the International Conference, Systems and Family Therapy: Ethics, Epistemology, New Methods, in Paris, France. However, despite its promising title, it provides nothing concrete or systematic for making systems ethical [9].

Stafford Beer's viable system model (VSM) is specific to hierarchically structured systems and associates ethics with a specific level of the hierarchy (System 5) [10]. Although every ethical system can be mapped onto the VSM structure, so too can every unethical system. And rather like creating an ethics committee, assigning "ethics" to a particular level of the architecture is insufficient to make a system behave ethically, it does not explain *how* to make a system ethical. It just creates the illusion of having solved the problem, but the problem has not been solved; only delegated. Although applying VSM *might* result in an ethical system, it is not inevitable. By contrast, we expect reliable ethicalness to be an inevitable emergent property of the entire system—if and only if the system is ethically adequate.

An important early step was to realize that the good regulator theorem is ambiguous because a regulator that is good at regulating is not necessarily good in an ethical sense. To avoid this ambiguity, this paper uses the term "effective" for the first meaning, "ethical" for the second meaning, and only uses "good" when both meanings are intended. It is only by imposing precision in the use of terminology that it was possible to clarify the otherwise muddled thinking and isolate the essence of an ethical system.

To identify more necessary properties, a selection of ethical and unethical systems were considered, including autonomous vehicles, bank ATMs, various flavours of capitalism, central banks, corrupt politicians, dating systems, democracies, dictatorships, healthcare robots, a jury, law-abiding citizens, money laundering banks, product design processes, superintelligent machines, the U.S. Supreme Court system, vehicle exhaust emission test cheating corporations, and voting machines.

Considering these diverse systems helped identify some general characteristics, such as having ethical goals, laws, and the intelligence to understand the laws and make rational decisions. Some other characteristics only became apparent after looking for ways that evil actors (internal or external to the system) could subvert each system, such as by hacking, tampering, feeding the system false information, or by threatening, bribing, and blackmailing people who have influence on the system.

The analysis was exploratory and unstructured. In considering such a wide range of different types of systems, it was only necessary to reflect on each system's unique or special differences to identify any new aspects that had not already been identified. The first few requisite properties were found by considering an abstract regulator and then an autonomous vehicle. Each further system that was considered contributed its own points of special interest. For example, systems that control money include strict requirements for an audit trail and physical resistance to tampering, a jury might be threatened, a judge must cope with liars, a supreme court justice might be blackmailed, corrupt politicians need to make secret deals and obfuscate the source of their wealth and what they did to get it, most computer systems accept their inputs as truth without corroboration and can be hacked, robots must obey laws, and in a village, gossip keeps track of who has a bad reputation, but on the internet, criminal, violent, and abusive men can keep creating new profiles to find more victims, who have virtually zero chance of discovering, before it is too late, that they are replying to a psychopath.

Finally, a minimum set of additional properties were identified that would counter the entire set of potential vulnerabilities. In all, nine properties were identified that are necessary and sufficient to guarantee that a system will behave ethically. These nine requisites are integrated in the ethical regulator theorem (ERT), which can be used as a decision function, IsEthical, that can be applied systematically to categorize any regulated system as being ethically adequate, ethically inadequate, or ethically undecidable. A proof of the theorem is provided. Another result of ERT is a basis (known as the MakeEthical function) for systematically identifying improvements that are necessary for a given system to be made ethically adequate. The IsEthical and MakeEthical functions can be used to construct an ethical design process that can be retrofitted to enhance any existing formal design process, such as VSM.

Since ERT did not seem to fit into the existing cybernetics framework, a new framework was developed out of necessity. It uses the IsEthical function to distinguish between two types of superintelligent machines; those that are ethically adequate and those that are ethically inadequate. Together, the intelligence and ethics dimensions are used to identify four well-defined classes of systems. These four distinct classes can be appended to the existing two levels of first-order and second-order cybernetic systems to create a six-level framework for classifying cybernetic and superintelligent systems. An unexpected consequence of trying to categorize ERT was the realization that third-order cybernetics should be defined as "the cybernetics of ethical systems".

Since the ethical regulator theorem can be applied to any regulated system in any domain, and offers a new and systematic approach to making systems more ethical, the implications for making the world a better place are significant.

One result of the exploration of the proposed six-level framework is the identification of a race condition that results in either a cyberanthropic utopia or a cybermisanthropic dystopia. This dystopic threat is well known, however, by identifying the exact nature of the race condition, it becomes clear what strategy must be employed to try to avoid the possibility that superintelligent machines could lead humanity into a dystopic disaster.

Since it is imperative for humanity to avoid this existential threat, concrete actions are proposed, including a grand challenge to apply ERT to new and existing systems in all areas of society in what is characterized as a systemic ethical revolution. And because an important component of that revolution is psychological, 82 ethically inspiring quotes by twelve famous empaths from five continents are presented that demonstrate that ethics transcends science, politics, genders, nations, and religions, and is probably the only force that can unify humanity to work together for our greater good.

2. The Ethical Regulator Theorem

The ethical regulator theorem (ERT) claims that the following nine requisites are necessary and sufficient for a cybernetic regulator to be effective and ethical:

(1) Purpose expressed as unambiguously prioritized goals.
(2) Truth about the past and present.
(3) Variety of possible actions.
(4) Predictability of the future effects of actions.
(5) Intelligence to choose the best actions.
(6) Influence on the system being regulated.
(7) Ethics expressed as unambiguously prioritized rules.
(8) Integrity of all subsystems.
(9) Transparency of ethical behaviour.

Of these nine requisites, only the first six are necessary for a regulator to be effective. If a system does not need to be ethical, the three requisites ethics, integrity, and transparency are optional. Figure 1 and the following sections explain the requisites in more detail.

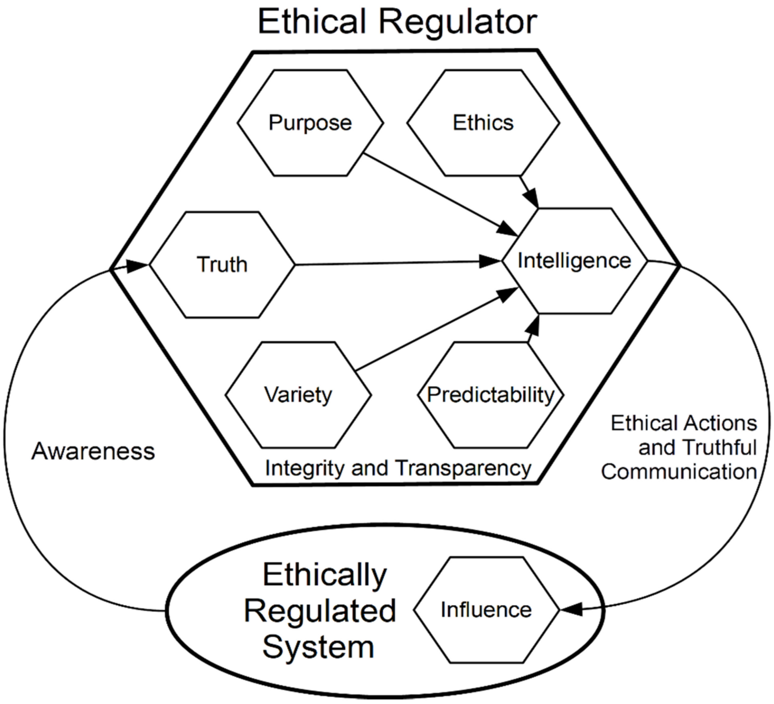

Ethical Regulator
Purpose
Ethics
Truth
Intelligence
Variety
Predictability
Integrity and Transparency
Awareness
Ethical Actions and Truthful Communication
Ethically Regulated System
Influence

Figure 1. An ethically regulated system.

2.1. Requisite Purpose

Because complex systems are required to satisfy multiple goals, purpose must be expressed as unambiguously prioritized goals. Without well-defined goals, the system cannot be effective and might randomly adopt or default to a goal that is regarded as unethical by the society that it exists in.

2.2. Requisite Truth

Truth is not just about information that the regulator treats as facts or receives as inputs, but also the reliability of any interpretations of such information. This is the regulator's awareness of the current situation, knowledge, and beliefs. If the regulator's information sources or interpretations are unreliable and cannot be error-corrected, then the integrity of the system is in danger. In extremis, if the perceptions of the regulator can be manipulated, it can be tricked into making decisions that are ineffective or unethical.

An ethical regulator does not require perfectly accurate information, but it must be sufficiently truth-seeking to be able to cope with uncertainties and minimize the impact of unreliable information, misinterpretations, and deliberate misinformation as best as it can. This is much like the requirement that a good judge (one that is both effective and ethical) must be able to reach reliable verdicts "beyond reasonable doubt" from unreliable evidence.

2.3. Requisite Variety

Variety in the range of possible actions to choose from must be as rich as the range of potential disturbances or situations. This is nothing other than the law of requisite variety.

2.4. Requisite Predictability

Predictability requires sufficiently accurate models of the system that is being regulated and of the regulator itself, to be able to rank the actions and strategies that will give the best outcome. This is nothing other than the good regulator theorem.

2.5. Requisite Intelligence

Intelligence is applied to the previous requisite types of information to select the most effective/ethical strategies and actions from the set of possible actions. And because the output of one regulator is generally an input to other regulators (systems or people), if the selected action is an act of communication, it must be as truthful as possible. Here, the name intelligence is used only to mean the ability to make an informed selection, and it should not be equated with human intelligence, though it can include it.

2.6. Requisite Influence

Influence is the existence of pathways to transmit the effects of the selected actions to the regulated system. This is not a property of the regulator, but a function of the connectivity relationships that span from the regulator's outputs to elements of the regulated system and its environment. If a regulator has no influence on the regulated system, it is not a true regulator, it is a simulation or passive observer, and there are no direct ethical consequences; which can be important when observing or simulating dangerous situations or testing a preproduction system.

Depending on the nature of the system that is being regulated, the speed and duration of the effects of actions can vary greatly. For example, a self-driving vehicle applying the brakes has an immediate effect on the vehicle's velocity, which lasts until the next acceleration; a new ruling by the Supreme Court has a much slower effect on society but could last for decades or possibly centuries; and the cascade that can be caused by someone sending a message to the network of transmission repeaters known as Twitter followers is unpredictably chaotic in both speed and duration.

In some systems, influence is more of a determining factor than variety. Indeed, the power of the law of requisite variety has often been overstated, for example, claiming that the subsystem with the most variety will control a system. ERT proves why this is not always true.

Let us consider two systems, A and B, that are competing to win control of system C, for example, two politicians seeking election. Often the variety of statements, actions, and strategies of the candidates is less important than their ability to purchase advertising to influence the voters.

And if a robber uses a gun to increase his effectiveness, the use of a gun does not amplify his variety, it is just one existing element in his range of possible variety, yet making that choice greatly increases his effectiveness at manipulating his victims. Such an increase in effectiveness, like buying advertising, is best explained in terms of an increase in influence.

In the light of the concept of influence, the belief that variety can be amplified appears to be as delusional as the idea that randomness can be amplified. Feeding variety or randomness into a genuinely noiseless amplifier cannot produce more variety or randomness than was fed into it. The variety of the robber or an advertising message is effectively constant.

The six requisites described so far are necessary and sufficient for a system to be effective but are not sufficient for it to be ethical.

2.7. Effectiveness Function

The ethical regulator theorem implies that we can define a function for the effectiveness that a regulator, R, has in controlling a system. It captures how the effectiveness of the regulator depends on the effectiveness of six requisites:

$$\text{Effectiveness}_R = \text{Purpose}_R \times \text{Truth}_R \times \text{Variety}_R \times \text{Predictability}_R \times \text{Intelligence}_R \times \text{Influence}_R \quad (1)$$

In this form, we would assign each requisite an effectiveness value between 0 and 1, where 1 means that it is perfect or optimal. And if the effectiveness of even just one of the requisites is close to zero, the effectiveness of the whole regulator is massively reduced. Applied to our two politicians: If $\text{Effectiveness}_A > \text{Effectiveness}_B$, then A is more likely than B to win control of system C.

However, it is neither necessary nor possible to calculate meaningful numerical values to compare the effectiveness of different systems or configurations. The essential value of the function is to understand the relationships and dependencies that it captures.

It is sufficient if an understanding of the effectiveness function informs the system design strategy; recognizing that a maximally effective system requires that the effectiveness of these six requisite dimensions are maximized, and that a successful attack on the integrity or effectiveness of any of them spells disaster for the effectiveness of the whole system.

It is worth noting that in social systems, money can buy media influence; and if the media is broadcasting lies, propaganda, or advertising, it reduces the quality of Truth_X that is received by every voter or consumer, X, which can manipulate them into making decisions that are not in their best interest.

2.8. Requisite Ethics

Ethics must be expressed as unambiguously prioritized laws, regulations, and rules that codify constraints and imperatives, for example, Isaac Asimov's First Law of Robotics: "A robot may not injure a human being or, through inaction, allow a human being to come to harm." [11], but ideally, expressed unambiguously in a formal language such as XML, which can be understood by humans and computers.

Ethical rules define constraints on the variety of actions and have a higher priority than the goals for purpose. By always obeying the relevant highest priority rules, the regulator is guaranteed to act ethically within the scope of the ethical schema, which provides a model of acceptable (ethical) behaviour. The ethical rules have the power of veto over possible actions and strategies, which makes it safe for AI to generate candidate strategies and actions algorithmically, without having to worry whether it might generate unethical possibilities.

Since ethical schemas vary between different cultures, in machines, they must be handled like plug-in software modules. And because an ethical schema can encode any ethics, good or evil, each ethical schema must be anchored explicitly in the laws of a legislative jurisdiction. When a person or system crosses a state or national border it is necessary to activate a different set of ethical schemas, i.e., a different set of laws, regulations, and rules. And the ethics modules must be prioritized so that it is unambiguous which module has precedence in the event of a conflict, for example, between national and state laws. The highest-level laws could be encoded in hardware to be unhackable.

A taxonomy of ethics modules can provide rules for all conceivable situations. For example, gun-law, traffic-rules, child-care, tax-law, contract-law, maritime-law, drone-flying, police-regulations, and warfare-rules-of-engagement. And the ethics modules can be treated like device drivers, so that to be fully operational, a hypothetical gun-carrying robot that can drive on roads requires valid ethics modules for gun-law and traffic-rules. Without the necessary ethics modules for the appropriate legal jurisdiction, the robot's gun or driving capabilities are automatically disabled.

By legislating that all autonomous artificial intelligence systems must obey appropriate ethics modules that are issued by an organization that is run by humans, we can establish a control mechanism that should ensure that intelligent machines are always subject to human ethics; without unduly restricting the freedom of AI researchers. In fact, it will free up AI researchers and knowledge engineers to focus on the more challenging requisites of truth, variety, predictability, and intelligence.

When we introduce ethics, the effectiveness function must be modified because the effect of behaving ethically is that it reduces the variety of options that are available, by removing all possibilities that are unethical. Thus, if A is an ethical politician and B is an unethical politician, we get something like the following:

$$\begin{gathered}\text{Effectiveness}_A = \text{Purpose}_A \times \text{Truth}_A \times (\text{Variety}_A - \text{Ethics}_A) \times \text{Predictability}_A \times \\ \text{Intelligence}_A \times \text{Influence}_A\end{gathered} \tag{2}$$

$$\text{Effectiveness}_B = \text{Purpose}_B \times \text{Truth}_B \times \text{Variety}_B \times \text{Predictability}_B \times \text{Intelligence}_B \times \text{Influence}_B \tag{3}$$

which captures the reality that politicians and businessmen who lie and cheat have an advantage over ones that are ethical.

2.9. Requisite Integrity

Integrity of the regulator and its subsystems must be assured through features such as resistance to tampering, intrusion detection, cryptographically authenticated ethics modules, and compliance with all laws, regulations, and rules. Monitoring mechanisms must detect if an invalid ethics module is being used or if an ethical constraint is violated, and if necessary, activate a fail-safe mode, preserve evidence, and notify the manufacturer and/or the appropriate authorities.

The regulator's first-order integrity mechanisms cannot protect the pathways on which the regulator depends to influence the system. This poses a potential vulnerability that can only be mitigated by using the awareness feedback to check for evidence of the effect of each action.

2.10. Requisite Transparency

Demanding to be trusted is unethical because it enables betrayal. Trustworthiness must always be provable through Transparency. So the law of ethical transparency is introduced, stating:

> For a system to be truly ethical, it must always be possible to prove retrospectively that it acted ethically with respect to the appropriate ethical schema.

Whereas it does not really matter whether the programmers of a chess playing robot can find out why a particular piece was sacrificed during a game, the logic of ethical decisions must never be hidden in the depths of opaque processes, neural networks, or lost to the passage of time. Generally, this requisite can be satisfied by having multiple independent witnesses or keeping an audit trail that is adequate and secure.

When an ethically adequate system violates an ethical constraint, as they sometimes will, analysis of the audit trail will identify the reason. For example, because a faulty neural network wrongly identified a boy leading a cow as a calf leading a man, or it will prove who in the chain-of-command knew what about illegal corporate activities.

Integrity and transparency are codependent security requisites: We require both integrity of transparency and transparency of integrity.

2.11. Evaluating Ethical Adequacy

Like a public health inspection of a restaurant, an evaluated system is judged on the adequacy of each requisite dimension. If and only if a system and all its significant subsystems satisfy all nine ERT requisites is it said to be "ethically adequate". Otherwise it is classified as "ethically inadequate" and the weaknesses listed with recommendations for improving them.

And because a truly ethical system must be maximally tamper-resistant and unhackable, the evaluation of ethical adequacy also has similarities to penetration testing and red teaming techniques; where the evaluation team tries to identify weaknesses and theoretical possibilities to subvert the integrity of the system and all its subsystems via all possible attack vectors and surfaces.

For each of the nine dimensions, D_i, the evaluators must consider the following three questions:

(a) How can the system fail or be subverted in the D_i dimension?

(b) How can the system be improved in the D_i dimension?

(c) Is the system adequate in the D_i dimension?

This requires that the system is considered in 27 different ways, which delivers a systematic evaluation of the system's strengths and weaknesses. This process is a significant improvement on ad hoc approaches that can give very different results, depending on who performs them.

The theorem cannot be used to certify that an ethical schema is ethical because schemas (laws, regulations, and rules) can vary arbitrarily between cultures. However, it can be used to help identify the root causes of crises and to evaluate the ethical adequacy of any proposed interventions [12]. In the near future, certified ethical consultants may specialize in auditing and certifying the ethical adequacy of existing and proposed, products, processes, laws, organizations, and systems.

3. Ethical Regulator Theorem: Proof and Consequences

Now that we understand the nine requisites better, is it possible to prove that they are indeed necessary and sufficient for a cybernetic regulator to be effective and ethical?

3.1. Proof of Necessity

Proving necessity is simple: One-by-one, for each of the nine requisites dimensions, D_i, ask yourself the question "Can a regulator be effective or ethical without D_i?"—If it cannot, then D_i is necessary. For example, "Can a regulator be effective or ethical without Truth?".

The answer in each case is rather obvious, especially if you refer to Figure 1 and, one-by-one, cover each requisite using your thumb, and then consider whether the resulting system can be effective or ethical without the obscured requisite. Table 1 summarizes the results, which confirm the necessity claims, including the claim that ethics, integrity, and transparency, are optional for systems that only need to be effective.

Table 1. Proof of necessity "by thumb".

Symbol	Requisite Dimension	Is Necessary to Be Effective [1]?	Is Necessary to Be Ethical?
D_1	Purpose	Yes	Yes
D_2	Truth	Yes	Yes
D_3	Variety	Yes	Yes
D_4	Predictability	Yes	Yes
D_5	Intelligence	Yes	Yes
D_6	Influence	Yes	Yes
D_7	Ethics	No	Yes
D_8	Integrity	No	Yes
D_9	Transparency	No	Yes

[1] For effectiveness, the positive results for necessity correspond to the solutions for $\text{Effectiveness}_R = 0$. I.e., when $\text{Purpose}_R \times \text{Truth}_R \times \text{Variety}_R \times \text{Predictability}_R \times \text{Intelligence}_R \times \text{Influence}_R = 0$, for example, when $\text{Truth}_R = 0$, but not when $\text{Transparency}_R = 0$. This agreement between the ethical regulator theorem (ERT) effectiveness function and Table 1 is unremarkable because the effectiveness function was constructed from the results of posing the same necessity question for each requisite. So the agreement does not confirm the correctness of the theorem, but by performing this exercise yourself, you can confirm the correctness of the effectiveness function and Table 1.

3.2. Proof of Sufficiency

Proving that the nine requisites are sufficient is not so simple. First, let us assert that in the real world, effective systems and ethical systems exist. Now, for all those such systems, do any of them rely on any information, ability, or other factors to achieve effectiveness or ethicalness that is not covered by the nine requisites?

It is claimed that for all systems that have been considered by the author, the answer is no. However, this claim is easily refutable because it will only take one person to find one example of a necessary factor that is not covered by the nine requisites to demolish the current claim of sufficiency. In the event of that happening, we would adapt the theorem, if necessary, adding another requisite, reassert the sufficiency claim, thank whoever found the missing requisite, and issue the challenge: "Okay, *now* find one!".

So, although it is impossible to prove that such an exception does not exist, we can assert that it will always be possible to extend the theorem to include any missing requisites that might be identified in the future, thus restoring the validity of the claim of sufficiency for all known systems that have been considered.

3.3. ERT Universality

Anyone who has the impression that ERT primarily applies to artificial intelligence, robots, self-driving vehicles, and autonomous weapons systems is urged to consider how the theorem can be applied to human systems that make decisions that affect people or the environment, such as organizations, corporations, education systems, government institutions, CEOs, or yourself.

Justice Stevens [13] provided an excellent example of identifying the ethical inadequacy of the "Citizens United" ruling: "The Court's ruling threatens to undermine the integrity of elected institutions across the Nation. The path it has taken to reach its outcome will, I fear, do damage to this institution.", which implies that there is a pressing need to evaluate the ethical adequacy of the entire U.S. Supreme Court system.

Since the ethical regulator theorem can be applied to any system that is required to make ethical decisions, the nine ERT dimensions define a domain-independent abstraction layer that can be used to map between any regulated systems. This creates a vocabulary, or isomorphism, that allows practitioners in one domain to communicate meaningfully with practitioners in seemingly unrelated domains, and share insights and solutions, for example, across artificial intelligence, corporate governance, education systems, and designing consumer products. Specialists in each domain can share their challenges and solutions to improving purpose, truth, variety, predictability, intelligence/strategy, influence, ethics, integrity, and transparency. For example, perhaps a cloud-based secure audit trail service that was developed for one specific domain can be used to help solve transparency and integrity in completely unrelated domains.

3.4. ERT Reflexivity and Algebra

If the ethical regulator theorem is genuinely universal, it must produce meaningful results for the following two special cases:

- When we apply ERT to itself.
- When we apply ERT to second-order cybernetics (2oC).

First, let us define a convenient algebra that allows us to express important assertions in this domain. We need to distinguish between: (I) the act of evaluating the ethical adequacy of a system and (II) the act of determining the set of transformations or interventions that are necessary to make a system ethically adequate:

(I) A decision function, IsEthical (S), returns the value True if system S is ethically adequate, it returns the value False if S is ethically inadequate, or it returns the value Undecidable if S is significantly inconsistent, contradictory, or opaque. The value Undecidable should be regarded as an error message rather than a type of system, however it is prudent to treat such systems as ethically inadequate until proven otherwise.

(II) A function, MakeEthical (S), returns a set of transformations or interventions to make system S ethically adequate. If S is already ethically adequate, the function returns an empty set, {}.

Now we can use this ERT algebra to make some interesting and controversial claims in Table 2.

Table 2. Some ERT algebra assertions.

C_n	Claim	Interpretation/Justification
C_1	IsEthical (ERT) = True	The ERT system fulfils all nine requisites of ERT and is therefore ethically adequate. It can only be used to make systems more ethical.
C_2	MakeEthical (ERT) = {}	The ERT system is sufficient to be ethically adequate. Nothing else is required.
C_3	IsEthical (2oC) = False	Second-order cybernetics is ethically inadequate. Unlike ERT, it has no intrinsic ethics or integrity, so it can be used to design good or evil systems. It doesn't go beyond achieving effectiveness.
C_4	MakeEthical (2oC) = ERT	To become ethically adequate, 2oC needs the set of ERT concepts.
C_5	IsEthical (2oC + ERT) = True	Nothing in 2oC is incompatible with ERT; which would cause the function to return Undecidable.
C_6	2oC + ERT = 3oC	Logically, the system that is created by joining the 2oC and ERT systems would be named third-order cybernetics (3oC).
C_7	MakeEthical (VSM) = ERT	ERT can be used to fix the ethical weaknesses of VSM.
C_8	IsEthical (Capitalism) = False	All flavours of capitalism are ethically inadequate.
C_9	MakeEthical (Capitalism) = {Ethics, Integrity, Transparency}	Capitalism might be adequate in the six requisites for effectiveness, but it is obviously deficient in Ethics (laws, regulations, and rules), Integrity (compliance), and Transparency (audit trails). These must all be increased to make capitalism more ethical.

3.5. *The Law of Inevitable Ethical Inadequacy*

We can build on the proof of necessity to derive this new law:

> If you don't specify that you require a secure ethical system, what you get is an insecure unethical system.

Most people have an intuitive understanding that this law is true, but ERT proves that when ethical adequacy is not specified as a requirement for a system design, the resulting design phase will, quite rightly, tend to optimize for effectiveness, and maximally avoid the extra costs that would be incurred by implementing the ethics, integrity, and transparency dimensions, which are optional for a system that only needs to be effective, thus guaranteeing that the resulting system is ethically inadequate and vulnerable to manipulation; by design.

4. Ethical Design Process

Figure 2 illustrates the elements of a design process, in which an analysis phase produces a requirements artefact, which is the input to the design phase that produces a specification artefact, which is used as the input to the implementation phase, which realizes the system.

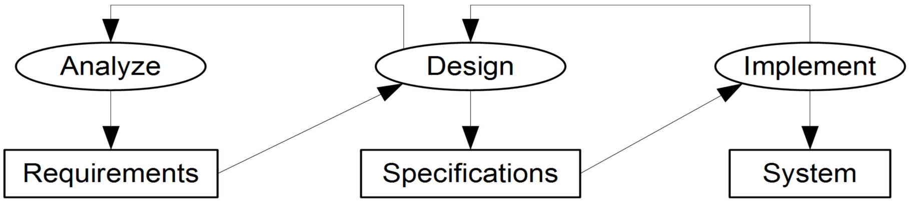


Figure 2. Ethically inadequate design process.

If a problem is found in the requirements during the design phase, feedback can trigger another iteration of the analysis phase. And if a problem in the specifications is found during implementation, feedback can trigger the design team to update the specifications or pass feedback to the analysis team to update the requirements.

Such a design process can be effective at producing systems that are effective, however, because the design process is ethically inadequate, it is inevitably only capable of reliably producing systems that are also ethically inadequate; and we cannot be sure that the resulting systems are not actually ethically evil, whether by accident or intentionally.

Fortunately, we can transform any effective but ethically inadequate design process, such as VSM, to make it ethically adequate by simply adding ethical adequacy acceptance testing of the requirements and specifications. How we can retrofit the ethical regulator theorem (ERT) to any effective design process that produces requirements and specifications before implementation starts is shown in Figure 3.

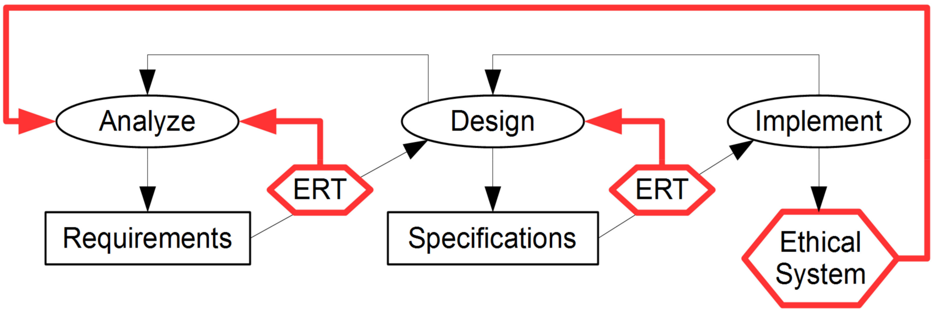


Figure 3. Ethically adequate design process.

To avoid inevitable ethical inadequacy, simply including ethical adequacy as a system requirement, will ensure that any effective but ethically inadequate design process that is employed must self-modify to include the ERT testing steps shown in Figure 3, otherwise the requirement cannot be fulfilled, and even basic acceptance testing will recognize the failure and reject the results.

Ideally, the ERT ethical adequacy evaluations should be performed by a team that includes all people who were involved in the production of the artefact that is being tested, who have deep internal knowledge of it, and certified ERT consultant-facilitator participants who are more objective, experienced, and have fewer blind spots. If an artefact is found to be ethically inadequate, it is rejected and recommendations for fixing the problems are provided as feedback to trigger another iteration of that phase. If an artefact is found to be ethically adequate, the artefact is accepted and passed onto the next phase. Since the two ERT testing steps ensure that the requirements and specifications are ethically adequate, if the implementation process performs an effective and lossless implementation of the specifications, the resulting system will also be ethically adequate.

This means that instances of the resulting system that are deployed in the real-world will include a real-time integrity monitoring mechanism that detects and reports any significant problems as feedback to the analysis team, which must decide whether the issue necessitates updating the system requirements, redesigning the specifications, a reimplementation, a remote configuration update, and/or activating an ethical fail-safe mode. Only if the system fails to enter its fail-safe mode might it be necessary for it to be deactivated using a kill switch or for it to be retired by a blade runner.

This concludes the description of the theorem and how to use it.

5. Discussion

The ethical regulator theorem has many far-reaching implications.

5.1. Legislative Implications

By creating a well-defined interface for coding ethics, it becomes easier to apportion liability for failures. For example, if a self-driving car crosses the border into India, fails to switch to the Indian government certified ethics module for traffic-rules, and in an emergency, decides to hit a cow to avoid hitting a dog, then the car manufacturer might be held liable for killing a sacred animal. But if the audit trail proves that the correct ethics module was activated, but the "don't hit cows" rule had an incorrectly low priority in the ethics schema, then the car manufacturer would not be liable.

It is foreseeable that one-day the laws and regulations of most countries will be published in a standardized computer-readable XML format, such as LKIF (legal knowledge interchange format), and cryptographically-signed by an official issuing authority. However, the existing governmental and regulatory organizations are inadequate to complete such a task in the necessary time frame. Perhaps, a non-profit organization without any conflicts of interests could define appropriate standards and start an open source ethics coding project for the laws, regulations, and rules that are most urgently required by the ethically adequate systems that we try to construct.

By standardizing ethics modules, systems from different manufacturers will all use identical ethics modules that are issued by national or international ethics authorities. The idea of having centralized ethics authorities might sound like part of a dystopic dictatorship but acting ethically is mostly just a matter of obeying laws, regulations, and rules, which are a normal and necessary part of every stable society. These ethics authorities could be independent of the legislative branch of government if the government lacks the necessary resources or commitment to unambiguous digital law-making.

Like Microsoft Windows operating system updates, when new laws, regulations, rules, or bug fixes to a previous ethics module are released, the new ethics module can be made available securely to all affected autonomous systems; crucially, including systems whose manufacturer has gone out of business or does not care about fixing end-user safety issues.

By comparison, Google's Android operating system provides a classic example of the law of inevitable ethical inadequacy. Google delegated the responsibility for issuing Android updates to the device manufacturers. And the inevitable and predictable consequence of that design decision is that most Android devices (87% in 2015) are insecure [14]. This exposes over one billion Android users to being hacked and their identity or financial details being stolen by criminals. The resulting chaos and the expensive suffering of the victims is not an innocent mistake, it is a typical effect of deliberately externalizing costs onto others and prioritizing a corporation's profits and competitive advantages over ethical consumer safety. Google could have designed it to ensure that Android security fixes are issued centrally, and are made available to all affected devices, even if the manufacturer has gone out of business. But if we cannot trust Google, who can we trust?

We certainly do not want robots, self-driving vehicles, and autonomous weapons systems relying on an update mechanism that stops working when the manufacturer goes out of business or decides to optimize its profits at the expense of security and safety updates.

Such unethical corporate behaviour must be legislated out of existence, otherwise it will keep recurring in different and damaging ways; causing unnecessary externalized costs, social chaos, and avoidable deaths. For example, ethically inadequate Internet-of-Things devices that send unencrypted data over the internet are vulnerable to being hacked, and will never receive security patches. Importing or selling such unethical devices that threaten our privacy and the security of our digital infrastructure must be made as illegal as it is to sell exploding cars, pharmaceuticals that contain lethal impurities, or passenger aircraft that automatically crash themselves.

5.2. Case Study: Boeing 737 MAX

On 29 October 2018, 13 min after takeoff, Lion Air Flight 610, a Boeing 737 MAX 8 aircraft, crashed [15]. It turned out that on the basic 737 MAX 8 model, the new manoeuvring characteristics augmentation system (MCAS) relied on a single angle of attack (AoA) sensor, and when the sensor failed, MCAS wrongly concluded that the plane was climbing too steeply, engaged improperly, and forced the aircraft's nose down, which caused it to crash into the sea at high speed, killing all 189 people on board. A week later, on 7 November 2018, the Federal Aviation Administration (FAA) issued an emergency airworthiness directive to make all airlines operating Boeing 737 MAX aircraft aware of new operating procedures relating to faulty AoA sensors.

Five months later, on 10 March 2019, six minutes after takeoff, Ethiopian Airlines Flight 302, a Boeing 737 MAX 8, crashed [16]. Again, the MCAS system had erroneously activated, and despite the pilot's efforts to prevent the resulting nose dive, the plane hit the ground at 1100 km/h, killing all 157 people on board. The next day, without any chance to investigate the second crash, the FAA asserted that the Boeing 737 MAX 8 model was airworthy. The day after that, Boeing CEO Dennis Muilenburg telephoned President Trump, and assured him that the model was safe. The day after that (13 March 2019) the FAA publicly admitted that there was a similarity between the two crashes, and grounded all Boeing 737 MAX aircraft. By 18 March, 387 aircraft operated by 59 airlines were grounded [17].

In such disasters, the regulatory body is often complicit due to either undue influence from the industry that they are supposed to be regulating (aka regulatory capture), or because of political pressure to avoid economic impacts. The FAA was certainly aware of the MCAS system's dependency on a single AoA sensor in November 2018, and if only it had taken that single point of failure seriously then, and grounded all 737 MAX aircraft in 2018, it would have prevented the 2019 Ethiopian Airlines Flight 302 crash, and saved 157 lives. It should not require a second crash for a known danger to be taken seriously.

Without doubt, IsEthical (Boeing) = False, but can we be confident that IsEthical (FAA) = True? Apparently, the FAA had a "longstanding delegation of regulatory authority to Boeing employees" [18], which creates obvious conflicts of interest, such as the Boeing employees feeling pressure not to cause production delays, increased costs, or reduced profitability. Perhaps industry regulators should be required to prove their honesty by being certified as ethically adequate. But who really regulates the regulators?

In July 2019, Boeing announced a $50 million fund to compensate the families of the 346 victims, which is equivalent to $144,500 per victim [19]. Six months later, Muilenburg was fired as CEO of Boeing, and was entitled to receive $62.2 million in compensation [20].

The grounding of the 737 MAX fleet has caused Boeing over $18 billion in losses [21]. And now that their financial problems have been made worse by the COVID crisis, it seems inevitable that Boeing will receive a bailout of tens of billions of dollars from the U.S. government to rescue them. But the cost of Muilenburg's avoidable 737 MAX disaster will not be paid for by Muilenburg, Boeing, or its shareholders, the politicians will externalize the bailout costs onto the U.S. tax payers.

We can identify the exact points in the ERT ethical adequacy evaluation process that the 737 MAX design weaknesses could have been identified. For example, if the MCAS system had been subjected to ethical adequacy evaluation, question a(D_2) asks "How can the MCAS system fail or be subverted in the truth dimension?".

Table 3 illustrates some of the problems that ERT could have identified long before the first crash and demonstrates the extensive coverage of the nine dimensions.

Any ethically adequate process would require further diligent investigation into all identified problems, require them to be fixed, and any decisions not to fix them should be documented thoroughly in a corporate audit trail that can be used in court to establish liability for negligence by the entire chain-of-command. This would create a strong incentive for all people involved (all the way up to CEO Muilenburg) to do the right thing during the design of the system. And the extra costs for Boeing to make their culture, processes, and products ethically adequate would have been a very small price to pay compared to 346 lives and 18,000,000,000 tax payer dollars.

Table 3. ERT ethical adequacy evaluation of the manoeuvring characteristics augmentation system (MCAS) system and its impact on the pilot.

Dimension	Problems That Could Have Been Identified Using ERT
Purpose	1. MCAS does not assist the pilot, it overrides the pilot.
Truth	2. Reliance on just one angle of attack sensor is a single point of failure. 3. Not making use of other available data to corroborate the sanity of the stall detection.
Variety	4. MCAS took variety away from the pilot. 5. Although turning off the electric stabilizer trim system inhibits MCAS, it also disables the electric trim switches in the cockpit, which leaves just the manual trim wheel in the cockpit as the only means to adjust the angle of the aircraft's horizontal trim stabilizer, which under certain scenarios, pilots might physically struggle to turn [22].
Predictability	6. If the software test system had been systematically evaluated, it would have identified that the test cases for the stall detection model had insufficient coverage, i.e., what happens if the single AoA sensor fails. 7. Boeing avoided the extra costs of retraining pilots on the new differences, such as between the 737 MAX and its predecessor model 737 NG. Inadequate training causes pilots to have inaccurate mental models, and the resulting lack of predictably reduces their ability to behave effectively in an emergency situation involving MCAS.
Intelligence	8. The hardcoded strategy was to override the experienced human pilot. 9. MCAS should disengage if the aircraft is in danger of colliding with the ground or sea. Automatically forcing the nose down is not optimal.
Influence	10. MCAS had too much influence on the angle of the aircraft's horizontal trim stabilizer, which took necessary influence from the pilot.
Ethics	11. Assuming that the stall detection is always correct, overriding the human pilot, and providing no simple way for the pilot to deactivate MCAS. 12. Boeing decided that safety features such as a second AoA sensor and an indicator of any disagreement between the sensors would be sold as profitable optional extras [23]. How could the FAA approve this decision?
Integrity	13. Having just one AoA sensor blatantly violates the basic aviation rule that requires redundancy for safety-critical subsystems in passenger aircraft. 14. Many Boeing employees must have had grave concerns, but the corporate culture and fear of consequences must have intimidated them into silence.
Transparency	15. Boeing neglected to inform the airlines about MCAS, it removed any references to MCAS from the operations manual that 737 MAX pilots relied on, and the FAA did not object to this lack of transparency [24].

5.3. Making Corporations More Ethical

A review of unethical corporate failures, such as Boeing 737 MAX, vehicle exhaust emissions test cheating, and the constant stream of crimes by too-big-to-be-convicted banks, shows that CEOs are so obsessed with cutting costs, maximizing profits, boosting the share price, deliberately creating layers of plausible deniability, and stuffing money into their own pockets that they ignore the other responsibilities of an ethical CEO, such as taking personal responsibility to ensure strict compliance with the law, avoiding disasters, and improving long-term viability. But there is no adequate incentive for CEOs to care about these things. If they are already multimillionaires, even a total collapse of the company would have no material effect on them or their families. The only effective deterrent for such people is a realistic threat of having to go to prison for a long time.

The fines that financial regulators impose on banks that commit serious crimes are less than they make from such crimes, and the CEOs never go to prison, so the fines are just a cost of doing business [25] and are utterly inadequate deterrents. Viewed systemically, such fines actually encourage banks to operate as organized criminal enterprises that enjoy no liability for their crimes. So banks can commit crimes and threaten to collapse the economy until we "bail them out" again, i.e., give them even more money on top of what they stole. Imagine what would happen if the only punishment for stealing something from a shop was having to give it back. The mere existence of such ineffective perverse incentive "punishments" is suggestive that one way or another, the regulators and key politicians are all paid off; no regular citizens agree that such fines are adequate. And it has been like this for decades.

The banks, Wall street, and lobbyists certainly reward politicians of all major political parties with very generous political donations and lobbying perks, often followed by suspiciously lucrative revolving door jobs, such as Tony Blair's £2 million per year job as "part-time adviser" to JP Morgan [26] or the $153 million that Bill and Hillary Clinton got in "speaking fees" [27], much of it from groups that had lobbied the government [28], in what creates the unfortunate appearance of being conflicts of interests, bribes, protection money, commissions, and/or backpay for services rendered.

One might think that it is time to legislate that large corporations must buy insurance to cover the full costs of insolvency, bailouts, negligence, or illegal activities. Anything less than that creates a moral hazard that encourages risk taking, confident in the knowledge that the costs of catastrophic failure will be externalized onto the tax payers. The new (higher) insurance premiums would reflect the true costs of the risks that corporations currently externalize onto others. However, given our arguably broken, corrupt, and captured political systems, this idea is a good example of a naïve utopia that could work in theory, but cannot realistically be reached from where we are now because the majority of politicians do not dare to do *anything* that their donors, powerful lobbyists, or potential future "employers" do not want, allowing them to effectively veto absolutely anything.

An alternative hope lies with the tenacity of insurance companies to find ways to avoid paying out. Given that ERT ethical adequacy evaluations could prevent many corporate disasters, insurance companies might start asking claimants on existing corporate liability insurance policies "Can you prove that you performed ethically adequate diligence?" and refuse to pay for disasters that can now be regarded as systematically avoidable. In a Darwinian economy, grossly negligent corporations should be destroyed. But even without such extreme consequences, the greed of the insurance industry could create the necessary incentives for other greedy cost-cutting industries to improve their ethical adequacy, thus transducing the self-interest of insurance corporations into a force that improves the behaviour of other corporations and increases the greater good as a side-effect of greed. It is a sad indictment of our dysfunctional democracies that the reason that this possible solution has a better chance of success is because it does not rely on politicians, who are not the solution, but are an integral part of the systemic ethical problems. But if we cannot trust our politicians, Google, banks, car manufacturers, Boeing, or the FAA to do the right thing, who can we trust?

5.4. Classification Framework

Now let us consider where the ethical regulator theorem fits into the existing cybernetics framework. One might assume that the theorem belongs in second-order cybernetics, however, in a 1990 conference plenary presentation, Heinz von Foerster (who made the distinction between first- and second-order cybernetics in 1974) implied that combining ethics and second-order cybernetics is not something that he would have suggested:

> "I am impressed by the ingenuity of the organizers who suggested to me the title of my presentation. They wanted me to address myself to 'Ethics and Second-Order Cybernetics'.

> To be honest, I would have never dared to propose such an outrageous title, but I must say that I am delighted that this title was chosen for me." [9]

Table 4 lists some of the cybernetic community's definitions of first- and second-order cybernetics, as summarized by Stuart Umpleby [29].

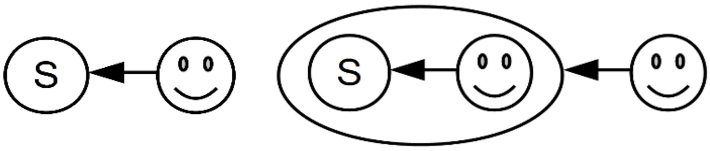
Table 4. Definitions of first- and second-order cybernetics.

Author	First-Order Cybernetics	Second-Order Cybernetics
von Foerster	The cybernetics of observed systems	The cybernetics of observing systems
Pask	The purpose of a model	The purpose of the modeller
Valera	Controlled systems	Autonomous system
Umpleby	Interaction among the variables in a system	Interaction between observer and observed
Umpleby	Theories of social systems	Theories of the interaction between ideas and society

Although every one of these definitions captures an important distinction, when compared to how the qualifiers "first-order" and "second-order" are used by other scientific communities, the cybernetic community's use of them appears to be rather subjective, lacks the consensus that is required by the scientific principle, and is of little utility, as required by Kuhn [30].

This incoherence in defining cybernetics as first-order and second-order not only prevents it from being useful to classify different types of systems and dissipates intellectual energy, but it also prevents the classification from being extended to higher orders, which can be viewed as either a self-limiting dead-end, or paradigmal autoapoptosis (self-programmed death), which is not entirely unlike the tragic situation of 39 members of the Heaven's Gate millennial death-cult, who believed that by committing suicide, they would be rescued by an alien spacecraft and "graduate to the Next Level" [31].

To illustrate the problem of classifying cybernetics into observer-centric "orders", let us start by considering first- and second-order cybernetics, as defined by von Foerster. Figure 4 illustrates how the observers' perspectives relate to a system, S.



(**a**) (**b**)

Figure 4. (**a**) First-order cybernetics. (**b**) Second-order cybernetics.

How can we use this paradigm to predict the future of cybernetics? Logically, third-order cybernetics would add a third observer's perspective, as shown in Figure 5.

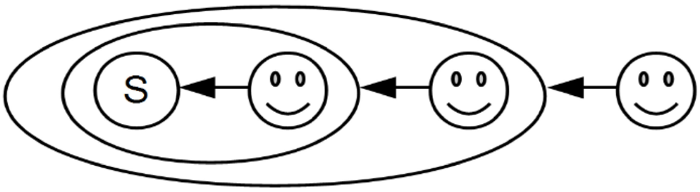


Figure 5. Third-order cybernetics.

However, from the perspective of the third observer, this looks more like psychology than cybernetics. In fact, this structure is isomorphic to a typical management team evaluation exercise, where the details of the task that is given to the team to work on is virtually irrelevant to the outermost observer. It can be any goal-oriented activity, such as building the highest stable tower possible from a limited set of Lego bricks, solving an impossible puzzle in a limited amount of time, or studying a first-order cybernetic system.

5.5. New Classification Framework

It could be of more utility to define "levels" of cybernetic systems that include categories of future systems that are already anticipated and associate each level with established concepts. To that end, Table 5 defines a six-level framework (6LF) for classifying cybernetic and superintelligent systems that makes use of the ERT IsEthical function to distinguish between two important subclasses of superintelligent systems.

Table 5. Six-level framework for classifying cybernetic and superintelligent systems.

Level [1]	The Cybernetics of	Also Known as	The Cybernetician
1	**Simple** systems	First-order cybernetics	Observes or uses the system
2	**Complex** systems	Second-order cybernetics	Participates in the system [2]
3	**Ethical** systems	Third-order cybernetics (or Cybernethics)	Designs the system
4	**Superintelligent** systems	Technological singularity	Stares incredulously, as the system redesigns itself
5	**Super-Ethical** systems (Superintelligent and ethically adequate)	Technological utopia or Cyberanthropic utopia	Is protected by the system
6	**Super-Unethical** systems (Superintelligent and ethically inadequate)	Technological dystopia or Cybermisanthropic dystopia	Is manipulated to obey the system

[1] Colours are assigned to the levels to help visually clarify their correspondence later in Figure 6.
[2] Klaus Krippendorff also highlighted the need to replace von Foerster's observer-centric orders, however, Krippendorff's concept of participation includes use, design, and conversation [32]. Here, some of von Foerster's distinction is preserved, so because a human participant in a system is regarded as being part of that system, and humans are complex, no system that includes human participation can be said to be simple.

Today, we are in the transition from building complex Cybernetic Level 2 systems to building ethical systems and superintelligent systems of Cybernetic Levels 3 and 4, and the future of our species and our fragile ecosystem is in our hands, but first, let us clarify each level and explore where this new framework leads us.

5.5.1. Cybernetic Level 1: Simple Systems

This level is about studying and designing simple systems that are effective. It is approximately equivalent to von Foerster's definition of first-order cybernetic systems.

5.5.2. Cybernetic Level 2: Complex Systems

This level is about studying and designing complex systems that are effective. It is approximately equivalent to von Foerster's definition of second-order cybernetic systems. There is still much important work to be done at this level.

It is important to emphasize that these definitions are different from von Foerster's definitions. According to his definition, a second-order cybernetic system always includes an observer. But complex systems may or may not include a human participant.

And although these two different definitions for Level 2 would often be in agreement about the classification level of any given system, there are some absurd edge cases that the new definition eliminates. For example, according to von Foerster's definition, the study of weather systems on Jupiter is first-order, but the study of Earth weather systems is second-order, because Earth-based meteorologists affect the Earth's weather system no less than the flapping wings of a butterfly. But under our definition of Level 2 systems as complex systems, studying weather systems falls into Level 2, regardless of the planet on which they occur.

Another important difference is that according to von Foerster himself, with his observer-centric definition, there is no possibility of there being a meaningful third-order cybernetics (see Section 5.9 Third-Order Cybernetics). By contrast, defining Level 2 systems as complex systems does not rule out the possibility of the existence of a Level 3, so in this respect, the new definition is not self-limiting.

It is acknowledged that this retuning of definitions might appear to make nonsense of the strict use of the terms first-order and second-order. So the use of these terms in the "Also known as" column in Table 5 for Levels 1 and 2 might now best be interpreted simply as an approximate mapping onto historical category names and not as scientific definitions. However, if we were to decide that the "orders" are now not levels of nested observers, but are orders of complexity or evolution, then the strict interpretation of applying the terms first-order, second-order, and third-order to categorize levels of systems is rigorously valid.

5.5.3. Cybernetic Level 3: Ethical Systems

In 1986, decades ahead of his time, it was the wonderful and inspiring Ranulph Glanville who defined "the cybernetics of ethics and the ethics of cybernetics" as "cybernethics" [33].

The ethical regulator theorem belongs at this level, which is concerned with designing human-made systems that are ethically adequate. These systems are new, and do not occur in nature. They must be designed to satisfy all nine requisites of the ethical regulator theorem. The regulating agents can be humans, machines, cyberanthropic hybrids, organizations, corporations, or government institutions. Ethically adequate autonomous machines must obey certified ethics modules.

In retrospect, now that we are not trying to extrapolate from just two points in concept-space, if Level 3 cybernetic systems are ethical, it is apparent that the third observer in the third-order cybernetics system of Figure 5 is not necessarily a psychologist or a lost cybernetician, but could be the second observer's conscience, her super-ego or higher-self, that constantly self-observing sense that we all have that knows the difference between right and wrong and between good and evil. This self-monitoring mechanism is known as integrity, and is something that today's ethically indifferent scientists, managers, executives, politicians, lawyers, bankers, and billionaires are woefully lacking. In non-psychopaths, integrity triggers feelings of bad conscience, regret, or guilt if it is ignored.

5.5.4. Cybernetic Level 4: Superintelligent Systems

The technological singularity is a hypothetical moment when a self-improvement process in a machine causes runaway improvements in intelligence that results in superintelligence that is far greater than any human mind. For this to happen, the system must be sufficiently self-aware of its own software and/or hardware specifications.

5.5.5. Superintelligence Tests

These types of self-awareness give rise to three levels of superintelligence abilities. The ability to reprogram better software for itself, redesign better hardware for itself, and the ability to do both.

Together with the Turing test [34], these tests mark milestones in the evolution of AI systems towards superintelligence and should cause us alarm if progress towards them is made without significant progress creating ethical systems first. Of these tests, the Turing test is the easiest to achieve because it is essentially a parlour game that only requires that a computer can imitate a (not necessarily very intelligent) human sufficiently well to convince humans most of the time that it is a human being and does not require self-awareness or runaway improvements in intelligence.

5.5.6. Prophecies of Possible Futures

In 1951, writing in his journal [35], Ross Ashby considered how to plan an advanced society as a "super brain" [36]. A year later, he described how super-clever machines could create a cyberanthropic utopia: "It may be found that we shall solve our social problems by directing machines that can deliver an intelligence that is not our own" [37].

Two pages later, he described a cybermisanthropic dystopia where a "Million I.Q. Engine" sounds like Facebook and Google, but on steroids: "What people could resist propaganda and blarney directed by an I.Q. of 1,000,000? It would get to know their secret wishes, their unconscious drives; it would use symbolic messages that they didn't understand consciously; it would play on their enthusiasms and hopes. They would be as children to it. (This sounds very much like Goebbels controlling the Germans)."

On the appearance of such a machine, he described a paradox of perception of higher intelligence: "It seems, therefore, that a super-clever machine will not look clever. It will look either deceptively simple or, more likely, merely random." [38]. On the same subject, Arthur C. Clarke's Third Law states: "Any sufficiently advanced technology is indistinguishable from magic." [39]. If you think that Clarke's "magic" and Ashby's "deceptively simple or merely random" are incompatible; take a moment to reflect on the magical simplicity and "randomness" of a Las Vegas magic show or Google's search results' pages.

Just as there are two diametrically opposite archetypes for genius; namely the benevolent good genius and the nasty evil genius, it is important not to conflate systems that are ethical with ones that are not ethical, by making them share the same name or category, such as "intelligent", "Christian", or "rich". To do so, misdirects our cognitive focus onto the attention-grabbing dimension and leads us into the temptation to ignore the most important dimension: good and evil.

5.5.7. Cybernetic Level 5: Super-Ethical Systems

The term "super-ethical" is proposed to refer to superintelligent systems that are ethically adequate. Of course, by the time that super-ethical systems exist, a friendlier name, such as "sapient", will have been adopted and the term "super-ethical" will seem quaintly archaic.

5.5.8. Cybernetic Level 6: Super-Unethical Systems

The term "super-unethical" is proposed to refer to superintelligent systems that are ethically inadequate. This term should always carry a certain stigma, like "weapons of mass destruction". No one who is working to create artificially intelligent systems should be allowed to escape admitting whether the systems are ethically inadequate.

Just as human genetic experimentation is strictly ethically regulated, we need legislation, regulation, standards, and certification to ensure that autonomous AI systems that make decisions that can have ethical consequences are subjected to the same kind of obsessively rigorous safety-oriented design, construction, and operating procedures as commercial aircraft, nuclear power stations, and vehicles that carry humans into space.

One could start arguing that intelligence is ethically neutral, and it is, but such arguments are fallacies because a hyper-genius "Million I.Q. Engine" without ethics is not ethically neutral. Even if it had ethical goals, it might break laws to achieve them. The possibility of creating a superintelligence that is ethically inadequate should be treated like a bomb that could destroy our planet. Even just planning to construct such a device is effectively conspiring to commit a crime against humanity.

As a thought experiment, let us imagine a hypothetical super-unethical version of Google, named the Googlevil Corporation. The CEO is Dr Evil, and both the CEO and the corporate AI are without ethics, avoid transparency, and will do anything to maximize their profits and power. The corporation's secret mission statement is "Collect and organize the world's personal information and make it accessible and useful for maximizing our profits, avoiding paying taxes, and blackmailing anyone in a position of power" and its secret corporate credo is "Sincerely say 'Believe me, we don't do evil', do it anyway, then look people in the eye and give them a creepy Zuckerberg-smile!".

Anytime that the hypothetical super-unethical Googlevil artificial intelligence or the imaginary psychopathic demagogue Dr Evil wants to blackmail the CEOs of other corporations, politicians that cannot be bought, jury members, or Supreme Court justices around the world to make "random" decisions that incrementally further their secret mission, would they have to do anything more than query the Googlevil user-profile database?

In theory, they would only need to be able to blackmail a majority of members of lower- and upper-houses (how hard can that be?) to be able to get any legislation that they want in any country or just a few Supreme Court justices to steer a nation into a fascist dystopia. By the time that super-unethical AI systems exist, they could be indistinguishable from their corporations, be immortal, immoral, and make unlimited donations (also known as bribes) to all Googlevil-friendly political parties in all techno-democratic dystopias on the planet. Does this already sound familiar?

5.6. Future Time-Line Bifurcation Race Condition

At this point in time, there is an existentially critical fork in our future time-line. Depending on whether the systems that achieve the singularity are ethically adequate or not, the runaway increase in intelligence and inevitable ethical polarization pressures will result in one of two outcomes:

- Good super-ethical AIs protect humanity and the biosphere.
- Evil super-unethical AIs dominate humanity and destroy the biosphere.

Figure 6 illustrates how plotting the ethical dimension orthogonally to the intelligence dimension clarifies the non-linear dependencies between different cybernetic levels, and clearly shows that the ethically inadequate superintelligent Level 4 systems have no dependency on us succeeding creating ethically adequate Level 3 systems first.

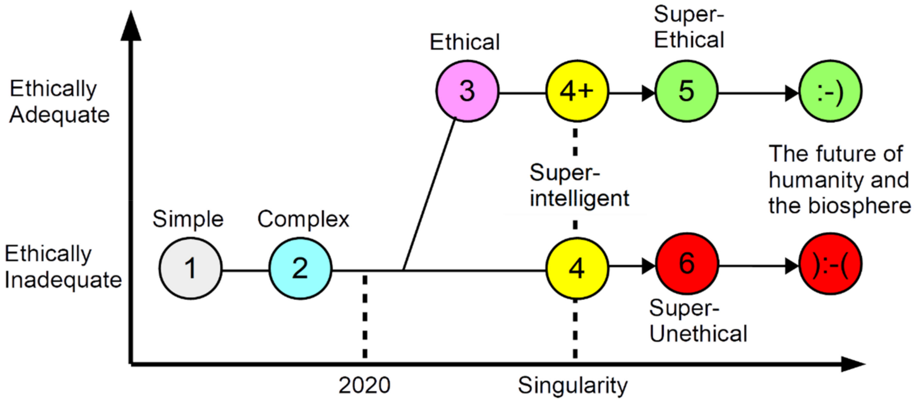


Figure 6. Two mutually exclusive possible futures.

If we continue on the current path from complex Level 2 systems to ethically inadequate superintelligent Level 4 systems, we will end up in a dystopia that is dominated by super-unethical Level 6 psychopathic systems, and the potential cyberanthropic utopia of being ruled by benevolent super-ethical Level 5 sapient systems will become permanently unreachable.

We must create ethically adequate Level 3 systems before we create superintelligent systems. This is the only way to ensure that they have ethical purposes, integrity, a synthetic sense of empathy, and are strictly law-abiding—before they become hyper-intelligent. Level 4+ systems will evolve into good super-ethical Level 5 sapient beings rather than evil super-unethical Level 6 psychopathic beings.

So, there is a race condition that will determine which of these two mutually exclusive possible futures will be the fate of our species; will our technological progress reach Level 4 or Level 4+ first? Level 4 and Level 6 systems must be made illegal. But will legislators regulate these technologies ethically and adequately, or will they sell us out for bribes or threats from special interest groups that will “campaign” for “self-regulation”? However, the horrific truth is that “self-regulation” by psychopathic CEOs actually means an unregulated race to create evil.

It cannot be overemphasized that the singularity (at 4/4+) is the point-of-no-return where humanity probably loses control over machines that are our intellectual superiors. And now is the tiny window of opportunity to ensure that they are programmed with ethics and purposes that serve the greater good of humanity and our fragile biosphere. In this context, it is clear that the ultimate grand challenge for third-order cyberneticians is to find ways to build ethical and super-ethical systems, avoid a cybermisanthropic dystopia, and help humanity create a super-ethical society.

5.7. Super-Ethical Society

Imagine how different the world would be:

- If we were happy to be ruled by wise and benevolent sapient beings that eliminated poverty, environmental destruction, corruption, and injustice.
- If the United Nations could deploy heavily armed super-ethical peace-keeping robot armies into conflict zones to protect civilians and enforce ceasefires.
- If we have super-ethical police robots that protect all citizens equally, 24 × 7, and never shoot our friends or family because of their race, religion, social class, lifestyle, or peaceful protesting.
- If super-ethical child-care robots accompany our children wherever they go, protecting them from danger, including physical, emotional, and sexual abuse.
- If ethically adequate corporations produce ethical products, provide ethical services, and pay ethical levels of unavoidable corporation tax.

Such a super-ethical society is possible; but only if we deliberately make it our goal, rise above polarizing politics, and act together in accordance with the undeniable truth that ethics is a higher power for good that transcends science, politics, genders, nations, and religions.

5.8. Cyberanthropic Utopia

Many ludicrous utopias have been proposed that are either just science fiction fantasies or are naïve designs that could work in theory, but can never realistically be reached peacefully from where we are now. So it is unsurprising that utopias have accumulated a bad reputation. But it is shockingly common for seemingly rational people to exhibit symptoms of classical conditioned-reflex (Pavlovian) negative responses to the stimulus word “utopia”; triggering emotional distress that disables their rational reasoning and evokes a childish response of ridicule. It is as if any serious use of the word “utopia” has become a reputation-threatening taboo. However, now that artificial intelligence is making such impressive progress and showing no signs of slowing down nor of having an upper-limit, Ross Ashby’s 68-year-old prediction looks increasingly realistic: We might be able to “... solve our social problems by directing machines that can deliver an intelligence that is not our own” [37].

In addition, Ashby’s prediction hints at a possible definition of a realistic minimum viable utopia:

> A world where our social problems have been solved.

Utopia need not mean a “perfect” society, or that we all have flying cars, robot servants, and never have to go to work. *Just fulfilling human needs and eliminating poverty would create a truly magnificent utopia*. And as we start making progress achieving it, many other human problems, such as starvation, malnutrition, parasitic diseases, homelessness, hopelessness-and poverty-driven prostitution, and crime will fade away and the world will become a very different and happier place to live in. Let no one say it cannot be done. Ethically adequate societies have existed in the past, where resources were shared and its members and the environment treated with respect. What is new is that we can now do it synthetically, consciously, deliberately.

5.9. Third-Order Cybernetics

Since Heinz von Foerster made the distinction between first- and second-order cybernetics in 1974, many people have attempted to find a plausible definition for third-order cybernetics, but no definition has gained acceptance. In a 1990 interview, he said " ... it would not create anything new, because by ascending into 'second-order,' as Aristotle would say, one has stepped into the circle that closes upon itself. One has stepped into the domain of concepts that apply to themselves." [9].

This paper proposes that third-order cybernetics should be defined as "the cybernetics of ethical systems", and that "the cybernetics of ethics", "Cybernetics 3.0", and "3oC" are all acceptable synonyms for it. Some of the supporting arguments for this proposal have already been mentioned, however a consolidated set of arguments are listed below:

1. Second-order cybernetics (2oC) discussions about the need to create ethical systems, including the need for cybernetics itself to embody ethics, did not produce any satisfactory solution. Here, "satisfactory solution" is understood to mean something like the ethically adequate design process of Figure 3, which can be used systematically to create real Level 3 systems that are ethically adequate. Recognizing this need but failing to fulfil the need could be referred to in the context of second-order cybernetics as "The Ethics Problem".
2. The fact that von Foerster described "Ethics and Second-Order Cybernetics" as "an outrageous title" strongly implies that ethics do not belong in the 2oC as he conceived it to be.
3. Whereas 2oC can be used for good or evil, the ethical regulator theorem can only be used for good. This is a fundamental difference that also implies ERT does not belong in 2oC. This claim to 3oC is not speculative: It is ERT's existence that now creates the need to define 3oC.
4. If we extrapolate from first-order cybernetics having one observer and 2oC having a second observer, we might expect 3oC to introduce a third observer. This hypothesized third observer maps exactly onto the ERT requirement that ethical systems must have real-time integrity mechanisms that monitor and enforce compliance of the system with respect to an appropriate ethical schema. And this third observer is different from the first and second observers.
5. Alternatively, we can deploy the good regulator theorem to derive a more rigorous justification: A first-order cybernetic regulator requires a model of the system being regulated, and a second-order cybernetic regulator can only achieve reflexivity by having a model of itself. Then to behave ethically, a cybernetic regulator needs a third model, a model of acceptable behaviour, which is encoded in the ethical schema. It is then simply a consequence of the fact that every model requires observations as inputs that brings into existence the need for three observing parts to exist in an ethical regulator, independently of whether a cybernetician is watching, or not. Echoing von Foerster's interview comments, Ranulph Glanville claimed that a third-order system cannot exist because it collapses into being equivalent to a first-order system [40], which might make sense in the observer-centric cybernetics paradigm, but it is nonsensical to suggest that the ethical regulator's third model (of acceptable behaviour) is equivalent to either its first model (of the system being regulated) or its second model (of itself). This demonstrates an important advantage of ERT's system-centric paradigm over the self-limiting observer-centric paradigm.
6. Ethically adequate systems are a rigorously defined and significant new type of system, and ERT plus the six-level framework (6LF) for classifying cybernetic and superintelligent systems (see Table 5) define a new branch of cybernetics that goes beyond achieving effectiveness, does not belong in 2oC, and provides an elegant solution to "The Ethics Problem". Logically, the system that is created by joining the 2oC and ERT systems would be named third-order cybernetics.
7. Although Glanville defined "the cybernetics of ethics and the ethics of cybernetics" as "cybernethics", this term is invented jargon that carries no meaning for people who are not familiar with its definition. By contrast, the term "third-order cybernetics" carries enough meaning for people who are familiar with the term "second-order cybernetics" to at least trigger interest

and curiosity. Therefore, using the term "third-order cybernetics" instead of "cybernethics" has advantages, and enhances 6LF by increasing symmetry in Table 5.

8. Together, ERT + 6LF create a new paradigm that has greater explanatory and predictive power than 2oC. For example, producing the ERT effectiveness function, the law of inevitable ethical inadequacy, identifying that the ethically adequate Level 3 systems are the missing type of cybernetic system that is necessary to integrate cybernetic systems and superintelligent systems into a common framework, explaining the impending bifurcation into either a cyberanthropic utopia or a cybermisanthropic dystopia (see Figure 6), and systematically identifying deficiencies in capitalism (see Table 2) and the Boeing 737 MAX (see Table 3). In addition, because 6LF integrates three classes of systems that do not yet exist, it can help us navigate a rational path into the future, for example, by predicting the existence of a race condition and thus identifying a possible solution to the dangers that are posed by superintelligent machines. Such insights cannot be obtained using 2oC.
9. Whereas it is impossible to define objectively which theories and practices belong in 2oC, making it an intimidating subject for outsiders to even contemplate mastering, ERT is defined and proved in eight pages and does not require knowledge of 2oC. This means that ERT, and how to apply it to any regulated system and in any domain, can easily be taught to non-cyberneticians, who will need no 2oC education. It is therefore logical and advantageous for ERT + 6LF to make a fresh start as third-order cybernetics, without being entangled with 46 years of unnecessary, fuzzy 2oC baggage. However, 3oC is not limited to ERT and 6LF, and will surely evolve before it matures.
10. Unlike 2oC, 3oC has a fundamentally ethical purpose (making systems ethical) that together with the proposed grand challenge and the vision of a super-ethical society, create a unique opportunity for the cybernetics and systems sciences community to take a leading role in implementing a long-overdue and much needed systemic ethical revolution. It is urgent that we develop this new field for governments, engineering, management, the other sciences, and all designers of systems to apply to the systems that they are responsible for. And this is a task that cannot be performed by any of the more narrowly defined branches of science.

If these arguments are accepted, it is suggested that going forwards, the term second-order cybernetics should only be used to refer to the second-order cybernetics of effectiveness without the ethical, integrity, and transparency aspects, which belong to the 3oC layer that can be used to transform any effective but ethically inadequate human-made complex Level 2 system, such as a design process, artificial intelligence, second-order cybernetics, or capitalism, into an ethical system.

Many people thought that cybernetics had faded away after peaking in the 1960s or early 1970s, but that peak was just a local maximum: Cybernetics is now rebooting as Cybernetics 3.0, and this time it will be more difficult to ignore because we will be applying requisite purpose, truth, variety, predictability, intelligence, influence, ethics, integrity, and transparency—and a tenth requisite, empathic love (Greek: Agape (https://en.wikipedia.org/wiki/Agape) [41]); because a sincere selfless desire to make the world a better place emerges only in people who love humanity and the biosphere unconditionally. Empaths care. We act according to our internal models of right and wrong [42], and do not actually need laws and threats of punishments for us to do the right thing. Basically, laws are made by bad people for bad people.

By contrast, deep down, non-empathic people (including most CEOs and politicians) really only care about themselves, and consequentially embody conflicts of interests that compel them to act against the greater good. Such non-empaths are not our allies, each and every one of them is part of the problem. Psychopaths simply do not care about the effects of their actions on others or the environment, regardless of the scale of the impact. They have a pathological internal model of what (for them) is acceptable behaviour that violates the principle of consensuality.

5.10. Our Future Epilog or Eulogy

We are rapidly approaching a perilous fork in the road in the evolution of complex Level 2 systems such as AI, democracies, immortal corporations, and human society, and it is urgent that we make these systems rigorously ethical before AI reaches the technological singularity, starts to evolve exponentially, exceeds human intelligence, and is used by ethically inadequate corporations, run by psychopaths, to dominate humanity politically and economically. Are they not doing this already?

We are the only generation that has the chance to steer the fate of future generations of humanity towards being collectively ruled, potentially for eternity, by benevolent super-ethical systems that create a stable cyberanthropic utopia for us, effectively and ethically minimizing human suffering and environmental problems. The alternative is to allow hubris, insatiable greed, and super-unethical systems to extinguish our rights and freedoms, and either enslave most of us in a cybermisanthropic dystopia or cause the extinction of our species to become a footnote in Gaia's geological record.

5.11. The Path Forwards

To start steering humanity and our wonderful planet towards becoming a stable cyberanthropic super-ethical society, we propose creating an independent, non-profit institute with ambitious goals that lie in the areas of research, development, standards, certification, legislation, and democracy.

5.11.1. Research and Development

The institute will promote theoretical and practical progress:

- Coordinate and fund research into creating ethical systems and making existing systems ethical.
- Develop a smartphone app to support teams perform rigorous ethical adequacy evaluations.
- Develop a taxonomy of open-source universal ethical schema modules for different types of laws, regulations, and rules that can be used by anyone, free of charge.

5.11.2. Standards and Certification

The institute will create an ethical certification infrastructure:

- Establish standards for certifying the ethical adequacy of systems.
- Establish a curriculum for training accredited ethical consultants.
- Coordinate and regulate contracts for ethical audits and certifications.

5.11.3. Legislation and Democracy

The institute will lobby governments to implement ethically adequate legislation and will evaluate the adequacy of any proposed legislation. In particular, promoting the following:

- Regulate autonomous machines to require that their design and implementation is ethically adequate, and that they support compulsory ethics modules.
- Require that all new systems, processes, and products are designed to be ethically adequate.
- Make it illegal to import or sell products that have not been certified as being ethically adequate, unless they are explicitly excluded from requiring certification.
- Extend political representation to every stakeholder in society by giving parents proxy votes to cast on-behalf of their children who are too young to vote, but not too young for morally bankrupt politicians to burden with unsustainable debt liabilities, while underfunding public education and health systems, and allowing unethical corporations to maximize short-term profits by devastating the environment for all future generations of humanity. What we currently call "universal suffrage" (https://en.wikipedia.org/wiki/Universal_suffrage) [43] is a shameless perversion of the true meaning of the word "universal".

5.12. Example: Applying ERT to Yourself

As members of a human society, we are all cybernetic regulators; of ourselves and of each other. As a thought experiment, to become a more effective and ethical force for good, you could identify ways to improve each requisite dimension as it applies to yourself. Table 6 provides an example of how you can use ERT to make yourself a better ethical regulator, or stated in ERT algebra: MakeEthical (yourself) = Table 6.

Finally, keep reviewing and refining your answers for purpose and ethics until they genuinely reflect who you are and how you want your world to become.

Table 6. Ways to become a better ethical regulator.

Requisite	Example Set of Self-Improvement Interventions
Purpose	To clarify your purpose in life and help you to recognize your strongest motivating thoughts, write down your most important life goals: 1. 2. 3. 4. 5.
Truth	To become a good judge (effective and ethical) of who tells the truth and who distorts it, seek alternative information sources that are genuinely independent of your primary sources. Investigate inconsistencies that you notice, modify the reputation of liars, and resolve to always doubt them sceptically in future.
Variety	Brainstorm new actions, responses, and strategies that you have never previously considered, to make progress towards achieving your goals.
Predictability	Improve your model of human behaviour by studying the following Wikipedia articles until you are competent at recognizing the patterns in yourself and others: • Cognitive biases (https://en.wikipedia.org/wiki/List_of_cognitive_biases) [44] • Defence mechanisms (https://en.wikipedia.org/wiki/Defence_mechanism) [45] • Fallacies (https://en.wikipedia.org/wiki/List_of_fallacies) [46] • Demagogue (https://en.wikipedia.org/wiki/Demagogue) [47]
Intelligence	Take a course or read a book on critical thinking or personal effectiveness.
Influence	Identify ways that you can increase your influence (on your family, friends, colleagues, or society) to achieve your life goals and promote your ethical values.
Ethics	Write down five undesirable, unethical, or disrespectful behaviours that, up until now, you have tolerated in corporations, organizations, or other people: 1. 2. 3. 4. 5. Next to them, write down five undesirable, unethical, or disrespectful behaviours that, up until now, you have tolerated in yourself. If you cannot think of five things about yourself, read the Wikipedia article: Denial (https://en.wikipedia.org/wiki/Denial) [48]. If that does not help, ask someone that you live with to suggest ten things that you do that they would prefer you not to do.
Integrity	Seek to stop or prevent all the undesirable, unethical, and disrespectful behaviours that you listed for ethics.
Transparency	Let other people know about the changes that you are making.

5.13. Ethically Resonant Wisdom

If you distil different solutions that contain alcohol, you get pure alcohol that is free of impurities. And if you distil different religions and philosophies that contain ethical wisdom, you get pure ethical wisdom that is free of culturally specific dogma. Such ethical wisdom is universal, and resonates with all good empaths, regardless of their worldview, political beliefs, gender identity, nationality, or religious affiliation. And because pure ethics is a higher power for good that transcends science, politics, genders, nations, and religions, it is probably the only force that can unify humanity to work together for our greater good.

For example, consider the following selected quotes:

Mahatma Gandhi (1869–1948):

(1) "The future depends on what you do today."
(2) "Be the change you wish to see in the world."
(3) "The difference between what we do and what we are capable of doing would suffice to solve most of the world's problems."
(4) "If I have the belief that I can do it, I shall surely acquire the capacity to do it even if I may not have it at the beginning."
(5) "First they ignore you, then they laugh at you, then they fight you, then you win."
(6) "Happiness is when what you think, what you say, and what you do are in harmony."
(7) "Non-cooperation with evil is as much a duty as is cooperation with good."
(8) "Poverty is the worst form of violence."
(9) "Capital as such is not evil; it is its wrong use that is evil."
(10) "There is sufficiency in the world for man's need, but not for man's greed."
(11) "There are people in the world so hungry, that God cannot appear to them except in the form of bread."
(12) "Those who say religion has nothing to do with politics do not know what religion is."
(13) "Where love is, there God is also."
(14) "God has no religion."
(15) "There is a higher court than the courts of justice and that is the court of conscience."
(16) "They may torture my body, break my bones, even kill me. Then they will have my dead body, but not my obedience."
(17) "Victory attained by violence is tantamount to a defeat, for it is momentary."
(18) "What difference does it make to the dead, the orphans, and the homeless, whether the mad destruction is wrought under the name of totalitarianism or the holy name of liberty or democracy?"
(19) "Your beliefs become your thoughts, your thoughts become your words, your words become your actions, your actions become your habits, your habits become your values, your values become your destiny."

His Holiness Pope Francis:

(20) "We must restore hope to young people, help the old, be open to the future, spread love. Be poor among the poor. We need to include the excluded and preach peace."
(21) "Hatred is not to be carried in the name of God. War is not to be waged in the name of God!"
(22) "Human rights are not only violated by terrorism, repression, or assassination, but also by unfair economic structures that create huge inequalities."
(23) "The worship of the golden calf of old has found a new and heartless image in the cult of money and the dictatorship of an economy which is faceless and lacking any truly human goal."

(24) "Men and women are sacrificed to the idols of profit and consumption: It is the 'culture of waste'. If a computer breaks it is a tragedy, but poverty, the needs and dramas of so many people end up being considered normal."

(25) "Women in the church are more important than bishops and priests."

(26) "All that is good, all that is true, all that is beautiful, God is the truth."

(27) "We all have the duty to do good."

(28) "Everyone has his own idea of good and evil and must choose to follow the good and fight evil as he conceives them. That would be enough to make the world a better place."

His Holiness the Dalai Lama XIV:

(29) "All religious institutions, despite different philosophical views, all have the same message—a message of love."

(30) "If you can, help others; if you cannot do that, at least do not harm them."

(31) "The whole purpose of religion is to facilitate love and compassion, patience, tolerance, humility, and forgiveness."

(32) "Irrespective of whether we are believers or agnostics, whether we believe in God or karma, moral ethics is a code which everyone is able to pursue."

(33) "The ultimate authority must always rest with the individual's own reason and critical analysis."

(34) "The true hero is one who conquers his own anger and hatred."

(35) "A good friend who points out mistakes and imperfections and rebukes evil is to be respected as if he reveals the secret of some hidden treasure."

(36) "A lack of transparency results in distrust and a deep sense of insecurity."

(37) "In our struggle for freedom, truth is the only weapon we possess."

(38) "Where ignorance is our master, there is no possibility of real peace."

(39) "Through violence, you may 'solve' one problem, but you sow the seeds for another."

(40) "Don't ever mistake my silence for ignorance, my calmness for acceptance or my kindness for weakness. Compassion and tolerance are not a sign of weakness, but a sign of strength."

(41) "A truly compassionate attitude toward others does not change even if they behave negatively or hurt you."

(42) "I defeat my enemies by making them my friends."

(43) "When you practice gratefulness, there is a sense of respect toward others."

(44) "With realization of one's own potential and self-confidence in one's abilities, one can build a better world."

(45) "If you think you are too small to make a difference, try sleeping with a mosquito."

(46) "As people alive today, we must consider future generations: A clean environment is a human right like any other. It is therefore part of our responsibility toward others to ensure that the world we pass on is as healthy, if not healthier, than we found it."

(47) "The ultimate source of happiness is not money and power, but warm-heartedness."

(48) "The more you are motivated by love, the more fearless and free your action will be."

(49) "Love and compassion are necessities, not luxuries. Without them humanity cannot survive."

(50) "Love is the absence of judgement."

(51) "Be kind when possible. It is always possible."

Dr Martin Luther King Jr. (1929–1968):

(52) "We must discover the power of love, the power, the redemptive power of love. And when we discover that we will be able to make of this old world a new world. We will be able to make men better. Love is the only way."

(53) "I say to you, 'I love you. I would rather die than hate you.' And I'm foolish enough to believe that through the power of this love, somewhere, men of the most recalcitrant bent will be transformed."

(54) "Darkness cannot drive out darkness; only light can do that. Hate cannot drive out hate, only love can do that."
(55) "Those who love peace must learn to organize as effectively as those who love war."
(56) "True peace is not merely the absence of tension. It is the presence of justice."
(57) "Injustice anywhere is a threat to justice everywhere."
(58) "In a real sense, all life is inter-related. All men are caught in an inescapable network of mutuality, tied in a single garment of destiny. Whatever affects one directly, affects all indirectly."
(59) "Every man must decide whether to walk in the light of creative altruism or in the darkness of destructive selfishness."
(60) "The time is always right to do the right thing."
(61) "We must learn that passively to accept an unjust system is to cooperate with that system, and thereby to become a participant in its evil."
(62) "You are not only responsible for what you say, but also for what you do not say."
(63) "Our lives begin to end the day we become silent about things that matter."
(64) "A nation that continues year after year to spend more money on military defense than on programs of social uplift is approaching spiritual doom."
(65) "Our scientific power has outrun our spiritual power. We have guided missiles and misguided men."
(66) "We should never forget that everything Adolf Hitler did in Germany was 'legal' and everything the Hungarian freedom fighters did in Hungary was 'illegal'."
(67) "Nonviolence is directed against forces of evil rather than against persons who happen to be doing evil. It is evil that the nonviolent resister seeks to defeat, not the persons victimized by evil."
(68) "Nonviolence means avoiding not only external physical violence but also internal violence of spirit. You not only refuse to shoot a man, but you refuse to hate him."

Nelson Mandela (1918–2013):

(69) "Freedom can never be taken for granted. Each generation must safeguard it and extend it. Your parents and elders sacrificed much so that you should have freedom without suffering what they did. Use this precious right to ensure that the darkness of the past never returns."
(70) "Like slavery and apartheid, poverty is not natural. It is man-made and it can be overcome and eradicated by the actions of human beings."
(71) "Overcoming poverty is not a gesture of charity. It is an act of justice."
(72) "As long as poverty, injustice and gross inequality persist in our world, none of us can truly rest."
(73) "Education is the most powerful weapon which you can use to change the world."
(74) "It is in your hands to create a better world for all who live in it."
(75) "May your choices reflect your hopes, not your fears."

Percy Bysshe Shelly (1792–1822):

(76) "Rise like lions after slumber"
In unvanquishable number!
Shake your chains to earth like dew
Which in sleep had fallen on you:
Ye are many—they are few!

What is Freedom?—ye can tell
That which slavery is, too well—
For its very name has grown
To an echo of your own.

'Tis to work and have such pay
As just keeps life from day to day
In your limbs, as in a cell
For the tyrants' use to dwell,

So that ye for them are made
Loom, and plough, and sword, and spade,
With or without your own will bent
To their defence and nourishment.

'Tis to see your children weak
With their mothers pine and peak,
When the winter winds are bleak,—
They are dying whilst I speak."

Norbert Wiener (1894–1964):

(77) "The hour is very late, and the choice of good and evil knocks at our door."

Stafford Beer (1926–2002):

(78) "The purpose of a system is what it does. There is after all, no point in claiming that the purpose of a system is to do what it constantly fails to do."

Albert Einstein (1879–1955):

(79) "No problem can be solved from the same level of consciousness that created it."

Margaret Mead (1901–1978):

(80) "Never doubt that a small group of thoughtful, committed citizens can change the world; indeed, it's the only thing that ever has."

Leonardo da Vinci (1452–1519):

(81) "I have been impressed with the urgency of doing. Knowing is not enough; we must apply. Being willing is not enough; we must do."

Bertolt Brecht (1898–1956):

(82) "Change the world: She needs it!"

Despite the authors of these quotes being separated by space, time, and their affiliations, it is easy to imagine that they all share the same empathic ethical belief system that respects all human and animal rights, the biosphere, consensuality, and the freedom to self-actualize to achieve our full potential. These empaths would surely have no significant arguments with each other if all twelve of them came together in one room to plan an ethical revolution to make the world a better place.

5.14. The Law of Evil Arguments

It is a certainty that all good empaths (without exception) are supportive of redesigning unethical systems, organizations, corporations, products, taxes, laws, regulations, and processes to make them more ethical. And it makes sense that the only people who want such systems to remain unethical and vulnerable to tampering and abuse are the small minority of people who benefit (directly or indirectly) from those systems remaining unethical.

The final law in this manifesto for a nonviolent global ethical revolution to create a stable cyberanthropic super-ethical society is the law of evil arguments, which asserts:

> Since no ethical argument can exist against making the world more ethical, anyone who argues against this goal, obstructs progress towards this goal, or abuses its sincere supporters, is objectively unethical or evil.

6. Conclusions

Ethically adequate systems are a rigorously defined and significant new type of system, and the ethical regulator theorem creates a theoretical basis that enables us to systematically evaluate, improve, and design ethically adequate systems. Since it is a universal theorem that can be applied to any regulated system, the possible areas of application are vast and potentially world-changing.

The six-level framework for classifying cybernetic and superintelligent systems leads to a theory-based solution to the danger that superintelligent machines might cause a cybermisanthropic dystopia: We must make all ethically inadequate Level 4 and Level 6 systems illegal, and strive to create ethically adequate Level 3 systems before we create superintelligent Level 4+ systems.

By creating a well-defined decision function (IsEthical) that identifies systems as being either ethically adequate or ethically inadequate, ERT provides a semantic precision that avoids the unstated assumptions and ambiguities that multiply exponentially when the word “ethical” is bandied around (generally accompanied by a lot of random hand-waving) as if we all understand it to mean the same thing. But “ethical AI”, “ethical product”, and “ethical corporation” mean very different things to different people. By contrast, ERT gives terms like “ethically adequate AI”, “ethically adequate product”, and “ethically adequate corporation” a much more precise meaning, and could even be made the subject of a formal certification process that qualifies recipients to use an ethically adequate branded logo and perhaps reduce their liability insurance premiums.

ERT’s universality means that the nine dimensions define an abstraction layer that can be mapped onto every regulated system in all domains, thus enabling communication and learning to take place between experts in seemingly unrelated fields.

Due to the flaw that was identified in Heinz von Foerster’s ethical imperative, a new definition is proposed, which is intended to embody both the essence of the proposed grand challenge and a principle for good that is universal and worthy of the delightfully magniloquent name, ethical imperative:

> Always strive to make new and existing systems ethically adequate.

The proposed grand challenge to implement a systemic ethical revolution is neither a new religion nor a political movement, it is a response to Johann Eder’s 2010 call for a grand challenge in Vienna [49] and Irma Wilson and Pamela Buckle Henning’s 2015 call to action for the systems sciences community in Berlin [50].

This ethical revolution is the product of a compassionate heart and mind, employing the ethical regulator theorem to generate maximally coherent ethical interventions in multiple complex Level 2 systems, such as the computational, corporate, criminal, cybernetic, economic, educational, environmental, personal [51], political, product development, psychological, regulatory, scientific, social, and spiritual [52] realms. And all such interventions not only resonate with each other in multiple dimensions, but also resonate with all good empaths who have ever existed or ever will.

This revolution is long overdue, and we are privileged to live in these exciting times, but passively watching from the sidelines, or doing nothing, only helps the criminals, psychopaths, demagogues, and ethically inadequate corporations to create and exploit the pathological chaos and emergent problems that, until now, we have accepted as normal. It is time for all good empaths to identify as good empaths and make a commitment to yourself, to each other, and to the forces of goodness, to do everything that you can to design, build, educate, organize, campaign, fight, heal, love, and pray for a better world.

> “To be bold enough to consciously and deliberately reach beyond ourselves, to accept a grand challenge for the greater good, would be an act of self-actualization.”—Stella Octangula [53].

We must demand strict legislation and higher standards to force ethically inadequate corporations to stop their races to the bottom, cost externalization strategies, deceitful denials of wrongdoing, and reckless plundering of the biosphere into extinction [54].

Just like we have regulations, legislation, and non-negotiable expectations that passenger aircraft are designed to include expensive redundant subsystems to avoid having single points of failure in flight-safety-critical systems, and that all electrical products that we purchase conform to strict safety standards, we must change our attitudes, to create a cultural shift that makes it totally unacceptable and utterly unthinkable to knowingly design systems or sell products that are ethically inadequate. Outrage at such behaviour is appropriate. Only ethically adequate corporations can be trusted to produce ethically adequate technologies, products, and services that make the world a better place for the entire human race and protect the biosphere from insatiable greed and destruction.

Arguably, the root cause of all evil is a lack of ethics, and by systematically applying the ethical regulator theorem, we can reliably increase ethical behaviour in many classes of systems (including AI, corporations, government, and industry regulators); progressively reducing unethical behaviour, unethical suffering, avoidable deaths, and social chaos, and increasing justice, happiness, well-being, and social order—finally setting humanity on a worthy journey to the tipping-point where we will experience a peaceful social phase-transition to a stable super-ethical utopic sapientocracy that is governed by wisdom [55], and as a consequence; salvation from evil.

Though this paper covers many topics, these are but means; the end has been throughout to make clear what principles must be followed when one attempts to restore ethical function to a sick organism that is, as a human society, of fearful complexity. It is my faith that the new understanding may lead to super-ethical systems that can create a better world, for the need is great.

Funding: This research received no external funding.

Acknowledgments: The author would like to express thanks to the following empaths for their inspiration by example: Vanilla Beer, Medea Benjamin, Luc Besson, Lee Camp, Noam Chomsky, CODEPINK, Adrian Cornish, Fabiana Crespo, Saskia Frisch, Sienna Grace, Matthew Marsden, Abby Martin, Lothar Merk, and Yvette Weis, and also sincere gratitude to the following empaths for their support, encouragement, helpful discussions, and/or constructive feedback on earlier versions of the paper: Nadisha-Marie Aliman, Zalina Baysarova, Kateryna Bieliaieva, Arend van Campen, Ranulph Glanville, Nick Green, Phillip Guddemi, Chris King, Andrea Lang, Allenna Leonard, Daniel Natal, Helen O'Shea, Patrick Pettit, Sherri Pritchett, Bernard Scott, Martina Sieber, Stuart Umpleby, Clément Vidal, Sarah Voskuhl, John Waters, Thomas Wilson, Stephen Wolfram, and the anonymous peer reviewers. Without their help, this paper would not exist. Thanks also to any readers who email the author with insights that are not covered in the paper.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK, 2014.
2. Pretel-Wilson, M. *Utopics: The Unification of Human Science*; Springer: Cham, Switzerland, 2020.
3. Conant, R.C.; Ashby, W.R. Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* **1970**, *1*, 89–97. [CrossRef]
4. Ashby, W.R. *An Introduction to Cybernetics*; Chapman and Hall: London, UK, 1956. Available online: http://rossashby.info/Ashby-Introduction-to-Cybernetics.pdf (accessed on 11 October 2020).
5. Cutler, A.; Pribić, M.; Humphrey, L. *Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers and Developers*; IBM Corporation: Endicott, NY, USA, 2019. Available online: https://ibm.biz/everydayethics (accessed on 11 October 2020).
6. European Commission. *High-Level Expert Group on Artificial Intelligence: Draft Ethics Guidelines for Trustworthy AI*; EC: Brussels, Belgium, 2018. Available online: https://www.euractiv.com/wp-content/uploads/sites/2/2018/12/AIHLEGDraftAIEthicsGuidelinespdf.pdf (accessed on 11 October 2020).
7. Metzinger, T. Ethics washing made in Europe. *Der Tagesspiegel* **2019**. Available online: https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html (accessed on 11 October 2020).
8. Von Foerster, H. On constructing a reality. In *Environmental Design and Research*; Preiser, F.E., Ed.; Dowden Hutchinson and Ross: Stroudsburg, PA, USA, 1973; Volume 2, pp. 35–46.
9. Von Foerster, H. Ethics and second-order cybernetics. In *Understanding Understanding: Essays on Cybernetics and Cognition*; Springer: New York, NY, USA, 2003. Available online: https://www.pangaro.com/hciiseminar2019/Heinz_von_Foerster-Ethics_and_Second-order_Cybernetics.pdf (accessed on 11 October 2020).

10. Beer, S. *Brain of the Firm*; John Wiley and Sons: New York, NY, USA, 1972.
11. Asimov, I. Runaround. In *Astounding Science Fiction*; Simon & Schuster: New York, NY, USA, 1942.
12. Ashby, M. How to apply the ethical regulator theorem to crises. In *A Systemic Vision of the Crises: From Optimization to Change Strategy? Proceedings of the 10th Congress of the European Union for Systemics, Brussels, Belgium, 15–17 October 2018*; EUS: Brussels, Belgium, 2018; Volume 8, pp. 53–58.
13. Legal Information Institute. Opinion of Stevens J. Appeal. Supreme Court of the United States. Cornell University Law School. 2010. Available online: https://www.law.cornell.edu/supct/html/08-205.ZX.html (accessed on 11 October 2020).
14. Thomas, D.R.; Beresford, A.R.; Rice, A. *Security Metrics for the Android Ecosystem*; University of Cambridge: Cambridge, UK, 2015. Available online: https://www.cl.cam.ac.uk/~{}drt24/papers/spsm-scoring.pdf (accessed on 11 October 2020).
15. Wikipedia. Lion Air Flight 610. Available online: https://wikipedia.org/wiki/Lion_Air_Flight_610 (accessed on 11 October 2020).
16. Wikipedia. Ethiopian Airlines Flight 302. Available online: https://wikipedia.org/wiki/Ethiopian_Airlines_Flight_302 (accessed on 11 October 2020).
17. Wikipedia. Boeing 737 MAX Groundings. Available online: https://wikipedia.org/wiki/Boeing_737_MAX_groundings (accessed on 11 October 2020).
18. Langewiesche, W. What really brought down the Boeing 737 Max? *The New York Times Magazine*, 18 September 2019. Available online: https://www.nytimes.com/2019/09/18/magazine/boeing-737-max-crashes.html (accessed on 11 October 2020).
19. Shepardson, D. Boeing to Make $50 Million in Payments to 737 MAX Crash Victims' Families. *Reuters*, 17 July 2019. Available online: https://www.reuters.com/article/us-ethiopia-airplane-boeing-idUSKCN1UC1O0 (accessed on 11 October 2020).
20. Gelles, D. Fired Boeing C.E.O. Muilenburg will get more than $60 million. *The New York Times*, 10 January 2020. Available online: https://www.nytimes.com/2020/01/10/business/boeing-dennis-muilenburg-severance.html (accessed on 11 October 2020).
21. Gelles, D. Boeing Expects 737 Max costs will surpass $18 billion. *The New York Times*, 29 January 2020. Available online: https://www.nytimes.com/2020/01/29/business/boeing-737-max-costs.html (accessed on 11 October 2020).
22. Hemmerdinger, J. Simulator Tests Demonstrate 737 Max Manual Trim Difficulties. *Flight Global*, 2 April 2020. Available online: https://www.flightglobal.com/safety/simulator-tests-demonstrate-737-max-manual-trim-difficulties/137651.article (accessed on 11 October 2020).
23. Tabuchi, H.; Gelles, D. Doomed Boeing jets lacked 2 safety features that company sold only as extras. *The New York Times*, 21 March 2019. Available online: https://www.nytimes.com/2019/03/21/business/boeing-safety-features-charge.html (accessed on 11 October 2020).
24. Cassidy, J. How Boeing and the F.A.A. created the 737 MAX catastrophe. *The New Yorker*, 17 September 2020. Available online: https://www.newyorker.com/news/our-columnists/how-boeing-and-the-faa-created-the-737-max-catastrophe (accessed on 11 October 2020).
25. Henning, P.J. Guilty pleas and heavy fines seem to be cost of business for Wall Street. *The New York Times*, 21 May 2015. Available online: https://www.nytimes.com/2015/05/21/business/dealbook/guilty-pleas-and-heavy-fines-seem-to-be-cost-of-business-for-wall-st.html (accessed on 11 October 2020).
26. Helm, T.; Waterfield, B. Tony Blair to earn £2m as JP Morgan adviser. *The Telegraph*, 11 January 2008. Available online: https://www.telegraph.co.uk/news/politics/labour/1575247/Tony-Blair-to-earn-2m-as-JP-Morgan-adviser.html (accessed on 11 October 2020).
27. Yoon, R. $153 Million in Bill and Hillary Clinton Speaking Fees, Documented. CNN. 2016. Available online: https://edition.cnn.com/2016/02/05/politics/hillary-clinton-bill-clinton-paid-speeches/index.html (accessed on 11 October 2020).
28. Cook, L. Here's who paid Hillary Clinton $22 million in speaking fees. *U.S. News*, 22 April 2016. Available online: https://www.usnews.com/news/articles/2016-04-22/heres-who-paid-hillary-clinton-22-million-in-speaking-fees (accessed on 11 October 2020).
29. Umpleby, S. What comes after second order cybernetics? *Cybern. Human Knowing* **2001**, *8*, 87–89.
30. Kuhn, T. *The Structure of Scientific Revolutions*; University of Chicago Press: Chicago, IL, USA, 1962.

31. Wikipedia. Heaven's Gate (Religious Group). Available online: https://wikipedia.org/wiki/Heaven\T1\textquoterights_Gate_(religious_group) (accessed on 11 October 2020).
32. Krippendorff, K. Cybernetics's reflexive turns. *Cybern. Human Knowing* **2008**, *15*, 173–184. Available online: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1135&context=asc_papers (accessed on 11 October 2020).
33. Glanville, R. The cybernetics of ethics and the ethics of cybernetics. In Proceedings of the 21st American Society Conference for Cybernetics, Virginia Beach, VA, USA, 19–23 February 1986.
34. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
35. Ashby, W.R. *Journal of W. Ross Ashby*; 1928–1972; British Library: London, UK, 2003; Volume 1, p. 7189. Available online: http://rossashby.info/journal (accessed on 11 October 2020).
36. Ashby, W.R. Advanced society planned as a super brain. In *Journal of W. Ross Ashby*; British Library: London, UK, 2003; Volume 14, p. 3528. Available online: http://rossashby.info/journal/page/3527.html (accessed on 11 October 2020).
37. Ashby, W.R. Power and I.Q. have many similar properties. In *Journal of W. Ross Ashby*; British Library: London, UK, 2003; Volume 16, pp. 4276–4278. Available online: http://rossashby.info/journal/page/4276.html (accessed on 11 October 2020).
38. Ashby, W.R. Appearance of a super-clever machine. In *Journal of W. Ross Ashby*; British Library: London, UK, 2003; Volume 16, pp. 4279–4280. Available online: http://rossashby.info/journal/page/4279.html (accessed on 11 October 2020).
39. Clarke, A.C. Hazards of prophecy: The failure of imagination. In *Profiles of the Future: An Inquiry into the Limits of the Future*; Bantam Books: New York, NY, USA, 1962.
40. Glanville, R. The purpose of second-order cybernetics. *Kybernetes* **2004**, *33*, 1379–1386. Available online: https://pdfs.semanticscholar.org/a96d/9ff94ab675a4788a05e1aadf1e142ba695bc.pdf (accessed on 11 October 2020). [CrossRef]
41. Wikipedia. Agape. Available online: https://en.wikipedia.org/wiki/Agape (accessed on 11 October 2020).
42. Natal, D. Ethics Mental Models. YouTube. 2020. Available online: https://www.youtube.com/watch?v=IzeGgjSv8kk (accessed on 11 October 2020).
43. Wikipedia. Universal Suffrage. Available online: https://wikipedia.org/wiki/Universal_suffrage (accessed on 11 October 2020).
44. Wikipedia. List of Cognitive Biases. Available online: https://wikipedia.org/wiki/List_of_cognitive_biases (accessed on 11 October 2020).
45. Wikipedia. Defence Mechanisms. Available online: https://wikipedia.org/wiki/Defence_mechanisms (accessed on 11 October 2020).
46. Wikipedia. List of Fallacies. Available online: https://wikipedia.org/wiki/List_of_fallacies (accessed on 11 October 2020).
47. Wikipedia. Demagogue. Available online: https://wikipedia.org/wiki/Demagogue (accessed on 11 October 2020).
48. Wikipedia. Denial. Available online: https://wikipedia.org/wiki/Denial (accessed on 11 October 2020).
49. Eder, J. Grand challenges for computer science research. In *Cybernetics and Systems*; Trappl, R., Ed.; Ross Ashby Memorial Lecture of the International Federation for Systems Research: Vienna, Austria, 2010; pp. 19–25.
50. Wilson, I.; Buckle Henning, P. A call to action for the systems sciences community. In Proceedings of the 59th Annual Meeting of the International Society for the Systems Sciences, Berlin, Germany, 2–7 August 2015; Volume 1. Available online: https://journals.isss.org/index.php/proceedings59th/article/viewFile/2609/836 (accessed on 11 October 2020).
51. Wilson, T. *Authentic Man School: A Practical Guide for Next-Level Living*; Timothy Cooper Clayton Publishing: Deer Lodge, MT, USA, 2019.
52. Wilson, T. The Ethical Regulator Theorem. YouTube. 2017. Available online: https://youtube.com/watch?v=NLhUajpMOI4 (accessed on 11 October 2020).
53. Octangula, S. Structure, Environment, Purpose, and a Grand Challenge for the ASC. 2011. Available online: https://asc-cybernetics.org/CofC/wp-content/uploads/2011/02/Structure-Environment-Purpose-and-a-Grand-Challenge-for-the-ASC-V2.0.pdf (accessed on 11 October 2020).
54. Johnstone, C. The Ecosystem Is Dying Because It Is Vastly More Profitable to Destroy the Ecosystem Than to Preserve It. Twitter. 2020. Available online: https://twitter.com/caitoz/status/1315057395547009024 (accessed on 11 October 2020).

55. Mobus, G. Question Everything, 2008–2011. Available online: https://faculty.washington.edu/gmobus/Background/seriesIndex.html (accessed on 5 December 2020).

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.