11-22-2016

On the origin of synthetic life: Attribution of output to a particular algorithm

Roman V. Yampolskiy
University of Louisville, roman.yampolskiy@louisville.edu

On the origin of synthetic life: attribution of output to a particular algorithm

To cite this article: Roman V Yampolskiy 2017 *Phys. Scr.* **92** 013002

View the article online for updates and enhancements.

Invited Comment

On the origin of synthetic life: attribution of output to a particular algorithm

Roman V Yampolskiy

Computer Engineering and Computer Science, University of Louisville, KY, USA

E-mail: roman.yampolskiy@louisville.edu

CrossMark

Abstract

With unprecedented advances in genetic engineering we are starting to see progressively more original examples of synthetic life. As such organisms become more common it is desirable to gain an ability to distinguish between natural and artificial life forms. In this paper, we address this challenge as a generalized version of Darwin's original problem, which he so brilliantly described in *On the Origin of Species*. After formalizing the problem of determining the samples' origin, we demonstrate that the problem is in fact unsolvable. In the general case, if computational resources of considered originator algorithms have not been limited and priors for such algorithms are known to be equal, both explanations are equality likely. Our results should attract attention of astrobiologists and scientists interested in developing a more complete theory of life, as well as of AI-Safety researchers.

Keywords: designometry, evolution, falsifiability, genetic engineering, GMO, synthetic life, robot evolution

1. Introduction

In 1859 Charles Darwin published his famous work—*On the Origin of Species*. In it, he provided a naturalistic explanation for the origins of fossilized and living biological samples collected in different regions of planet Earth. Before publication of Darwin's theory of natural selection (currently integrated into what is known as the theory of evolution), the prevailing theory used to explain such samples attributed their origins to a supernatural cause commonly assumed to be God (s). Darwin's theory quickly became the dominant one accepted by majority of scientists as the best explanation for the origins of different species. Evolutionary theory has only consolidated its position over the years due to strong additional evidence from such diverse fields as genetics, anthropology, and computer science [42].

In particular, research in genetics, which was not available during Darwin's life, has provided a treasure trove of experiments used to confirm Darwin's theory. At the same time, recent unprecedented advances in genetic engineering [101], directed evolution [54, 82], reprogramming [107] and synthetic genomics [38, 81] have allowed scientists to create

genetically modified organisms (GMOs) [53], expand genetic code [69, 111], create synthetic DNA [40] and synthetic life [51, 83] and consider creation of synthetic human genomes [35]. With the development of the latest tool for genetic manipulation (CRISPR [38]) no fundamental limits remain to the engineering of novel synthetic life forms. With fields like evolutionary robotics [68, 80], artificial life [61, 91, 106, 113] and evolutionary computation [8] providing theoretical and experimental support for the creation of evolvable synthetic life it is worthwhile to think about the future directions in post-Darwinian evolutionary theory [16].

A major challenge we are likely to face in the near future is being able to tell synthetic life forms from natural ones. We are already experiencing a need to identify GMOs for proper labeling and compliance, with some early work reported in that domain [2, 33, 77, 88]. With advances in space exploration, particularly with spacecraft visiting moons and planets of the solar system a possibility of bacterial contamination of space objects by organisms from Earth becomes a real possibility. If such organisms are later rediscovered we would need to be able to determine their origin. Likewise, spacecraft returning from a mission may bring unknown

1

organisms to earth, despite our best precautions [5], again presenting us with what, for want of a better expression, we call the sample attribution problem. There is also a possibility of discovering extraterrestrial life, but we will not concentrate on this situation.

We can now setup an artificial environment in which the samples' origin and distribution is known in advance (unlike in Darwin's original problem) and attempt to select the correct explanation between modern evolutionary [73] and non-evolutionary theories [15, 25] in a side-by-side test, some-thing we were previously not able to accomplish. Can science accurately distinguish between naturally evolved and geneti-cally engineered life forms if the distribution at the outset is known? This would be easy to set up in the lab once we have access to a large number of synthetic life forms. We could, for instance, take the special case where the distributions are equal initially by placing 50 naturally evolved organisms (class A) and a 50 engineered organisms (class B) into a lab setting and challenge the scientific community to accurately attribute each sample with respect to class A or B. With individual samples represented by standalone artifacts, not historical records of multiple related samples. As a thought experiment we can imagine setting this up on another planet as a challenge for alien scientists/explorers to reduce the impact of knowing something about natural organisms on Earth. This produces a very clean and decisive experiment as our artificial setup removes any bias associated with results directly affecting ourselves as people on Earth and allows us to perform an experiment, the results of which can be eval-uated against known truth-values. This would give us a chance to evaluate our theories of origin of biological samples.

Occam's Razor [13], which states that among multiple possible hypotheses the simpler one should be selected, is typically used to argue that evolutionary theory provides a superior explanation to theories which may include an engi-neer as such theories have to also explain nature and origins of the said engineer resulting in a more complicated hypothesis. However, in our proposed experiment, the nature of the engineer is known and samples are chosen to have equal likelihood of being generated by evolutionary or synthetic means making application of Occam's razor erroneous.

2. Generalized sample attribution problem (GSAP)

The proposed problem of discerning synthetic life from naturally evolved life forms can be seen as a special case of the general problem of selecting the algorithm responsible for generating observed samples from a number of possible algorithms, in contrast to the original problem faced by Darwin of developing a naturalistic algorithm, which could be used to explain collected biological samples. This problem is a subset of Solomonoff Induction [104, 105] and science in general [89]. Given a set of observations, determine which of many theories best accounts for what was observed and accurately predicts future observations.

To distinguish it from Darwin's original problem we call this problem the GSAP or generalized Darwin's problem. GSAP can be expressed as a computer science problem, in terms of algorithms and digital data. Any type of scientific samples and DNA code in particular can be represented as a bit string. Algorithms capable of generating bit strings encoding collected samples can be subdivided into two main types: evolutionary algorithms (genetic algorithms, genetic programming, etc) and engineered algorithms (expert sys-tems, cognitive systems, etc). Hybrid types, such as algo-rithms engineered to evolve [63] and those, which evolve capability to do engineering are also possible. In the biolo-gical domain, such mixed types can also be a result of crossbreeding between genetically engineered and naturally occurring organisms.

For the purposes of our work, it is important to establish clear criteria that determine whether something is artificial or natural, as many samples will combine properties of both. Well-engineered designs are capable of adaptation and some evolved systems are capable of engineering. For example, Shapiro argues that we observe natural genetic engineering in evolution: '… much of genome change in evolution results from a genetic engineering process utilizing the biochemical systems for mobilizing and reorganizing DNA structures present in living cells' [97]. We will define engineered sam-ples as those which include any contributions from an intentional agent such as a human engineer, a definition which excludes natural evolution which is intelligent [112] and powerful, but not purposeful or intentional [48], optimization process [62].

Finally, the possibility remains that a third type of algorithm, one outputting random bits will also hit the target string[1] [124], but as the size of the bit string grows expo-nentially increasing computational resources would be required for this algorithm. Random algorithms could corre-spond to appearance of living forms by chance in some parts of the multiverse due to availability of necessary probabilistic resources [29]. It will happen if the Everett's many-worlds interpretation of quantum physics [36] is true or if an algo-rithm is used to generate every possible universe [94] leading to generation of all conceivable strings in some universe, but as we are looking for a generic procedure to evaluate samples from particular universes, random algorithms can be safely ignored.

3. Distinguishing naturally evolved life from engineered life

Analyzing properties of a particular evolutionary algorithm may allow us to discover features which can be used to dis-tinguish between engineered and evolved organisms. For one, we know that evolution takes a very long time to work so if we learned that only a limited amount of time was available for the formation of a complex sample that would indicate

---

[1] In computer science a string is a finite sequence of characters, a concept not related to the string theory in physics.

that it was not a product of natural evolution. Also, some features have not been found in natural systems and so their inclusion may indicate that engineering took place. For example, Minsky wrote: 'many computers maintain unused copies of their most critical 'system' programs, and routinely check their integrity. However, no animals have evolved like schemes, presumably because such algorithms cannot develop through natural selection. The trouble is that error correction then would stop mutation—which would ultimately slow the rate of evolution of an animal's descendants so much that they would be unable to adapt to changes in their environments' [75].

Many respected scientists discuss the apparent difficulty in distinguishing between natural and engineered systems. For example, Shapiro says: 'it is very important to recognize that living cells resemble man-made systems for information processing and communication in their use of mechanisms for error detection and correction.' [97]. Similarly, Dawkins says: 'biology is the study of complicated things that give the appearance of having been designed for a purpose' [28] and continues 'we may say that a living body or organ is well designed if it has attributes that an intelligent and knowledgeable engineer might have built into it in order to achieve some sensible purpose… any engineer can recognize an object that has been designed, even poorly designed, for a purpose, and he can usually work out what that purpose is just by looking at the structure of the object' [28]. More generally, Minsky addresses the need to change our thinking regarding teleological explanations: 'we now can design systems based on new kinds of 'unnatural selection' that can exploit explicit plans and goals, and can also exploit the inheritance of acquired characteristics. It took a century for evolutionists to train themselves to avoid such ideas—biologists call them 'teleological' and 'Lamarckian'—but now we may have to change those rules!' [75]. Because evolution is a powerful optimization process it is capable of producing designs (springs [99], gears [19], compasses [87], Boolean logic networks [90], digital codes [50] etc) which are just as complex as those produced by intelligent agents, meaning that any test designed for detecting intelligence via examination of artifacts will fail to determine the causal source [76].

The difficulty in identifying how any one particular sample originated is exacerbated by the fact that most observed evidence is equally likely to support either synthetic or natural origins hypothesis. Compare observations of certain properties in naturally evolved biological organisms with similar observations from engineered organisms or software: DNA similarities between organisms indicate that later samples evolved from earlier ones (e.g. homo sapiens evolved from homo erectus), but code similarities between different releases of a software project indicate that some code was reused (e.g. Windows NT and Windows XP). Poor design in nature can be explained by the fact that the evolutionary process has no foresight (e.g. blind spot in human eyes due to the location of nerve fibers in front of the retina), but poor design in engineered systems can be explained by the incompetence of the engineer (ex. Toyota brake problems). Vestigial organs in some animals (ex. the wings of flightless

birds) are well explained by deducing that the species is in the process of adapting to a changed environment, but in the world of engineering, outdated features are frequently observed because it may be costly to redesign the system to remove them (e.g. ashtrays on airplanes) or to keep the system backwards compatible (able to be used with an older piece of hardware/software). Animals evolved the ability to adapt to changing environments (e.g. seasonal fur change), but software is frequently designed to be adaptable to user preferences (e.g. Netflix learning what movies you like).

Similar analysis can be applied to other evidence frequently used to justify attribution of samples to only a single hypothesis. It is important to note that this dual explanation for evidence is symmetric, so 'classical' evidence of engineering has a well-fitting explanation in naturalistic evolution and vice versa.

3.1. GMO detection methods

In order to comply with recent GMO regulations a number of techniques have been proposed to identify GMOs [2, 33, 88]. Many protein and nucleic acid-based detection methods have been developed and used for identification and quantification of GMOs. Such techniques typically rely on direct matching of samples to available reference materials stored in databases of known GMOs, which might include sequence information of exogenous inserts as well as endogenous reference genes [31]. Such methods of direct matching do not work for undisclosed modifications.

3.2. Unevolvable elements (UE)

An interesting sub-problem in the forensic investigation of origins of biological samples is the study of UE. Such elements are components of the sample that could not arise via an evolutionary process. This happens because all precursor elements do not improve, and sometimes even lower, fitness of the organism, preventing the module (code fragment) in question from arising. We distinguish two types of such elements: type A, which decode to a meaningful plaintext and are too long to happen by chance and Type B, which represent narrow targets in the space of possible solutions, surrounded by broad moats of negative fitness. While UE of the first type are well documented [9, 10], existence in the real-world of elements of the second type remains an open question. Let us examine each type of UE, and review some examples of each.

In many cases, genetic engineers behind the project have no reason to hide their contribution and in fact may be interested in making sure that the organism is labeled in such a way that it is obviously seen as synthetic, for example with watermarks [39]. Labeling is also useful to make it possible to trace an organism's descendants to the originator, which can be very important, for example in case of patent disputes [49]. Such labeling may take a form of a digital signature, or plain text metadata such as inserting the text 'Made in USA'.

Meaningful [124] text encoded in DNA on purpose or left there by mistake during the design process (as comments

or inactive code) could be detected and extracted [10]. In fact, over the last few decades scientists have inserted text messages into natural living organisms [115], GMOs [47] and synthetic life forms [39]. Such messages range in length from a few symbols (such as 'E = MC2' [115]) to full text of books [23] with a complete archival systems in the works [14]. The actual encoding and decoding process is beyond the scope of this paper, but the interested reader is advised to read a survey of the topic by Beck *et al* [10]. As long as the length of the discovered message is not trivial an investigator can conclude that engineering took place and the organism is not 100% natural. Efforts to find such text [78, 135] preceded the ability of scientists to insert such messages. A search for signs of engineering in biological (genomic) information of any unattributed biological sample is just as reasonable as the SETI search of astronomical data. In particular with any samples acquired from extraterrestrial sources such biological SETI[2] [98] should be a recommended first step.

The main challenge comes from recognizing text as 'meaningful' particularly in cases of non-human engineers. Many attempts have been made to formalize 'meaningful' to represent 'the value of a message as the amount of mathematical or other work plausibly done by its originator' [11]. Different measures have been proposed under such names as 'potential' (Adleman [1]), 'incomplete sequence' (Levin and V'jugin [65]), 'hitting time' (Levin [66]), 'sophistication' (Koppel [58]), and 'intelligence based complexity' (Yampolskiy [124]) for what is best known as the logical depth [11] of a string. Bennet describes this concept as follows: 'Of course, the receiver of a message does not know exactly how it originated; it might even have been produced by coin tossing. However, the receiver of an obviously non-random message, such as the first million bits of pi, would reject this 'null' hypothesis, on the grounds that it entails nearly a million bits worth of ad-hoc assumptions, and would favor an alternative hypothesis that the message originated from some mechanism for computing pi. The plausible work involved in creating a message, then, is the amount of work required to derive it from a hypothetical cause involving no unnecessary, ad-hoc assumptions. It is this notion of the message value that depth attempts to formalize' [11]. Similarly, Gurevich, describes a step-by-step process for what he calls 'impugning randomness' [45] a method for distinguishing purposeful from accidental.

By analogy with the SETI approach a search for artificiality and cognitive universals could take place instead, with statistical abnormalities and non-randomness being used to detect language-like patterns [34, 64]. To avoid ambiguity it is desirable to find patterns which are (1) highly statistically significant and (2) exhibiting hallmarks of artificiality such as 'symbol of zero, the privileged decimal syntax and semantical symmetries' (3) inconsistent in principal with any natural process be it Darwinian or Lamarckian evolution [98].

In adversarial scenarios, such as illegal utilization of GMOs, genetic engineers might be interested in hiding their contribution to the design of the organism, either by explicitly

---

² Search for messages in biological information.

erasing all evidence or at least by making its detection difficult if not impossible without privileged information by relying on steganography [57] or deniable cryptography [92, 133]. Deniable cryptography produced by encoding and combination of multiple plain texts is not very efficient in terms of size of cipher text and would produce large segments of DNA with no discernable meaning, something akin to 'junk DNA'. However, recent research suggests that such DNA segments are actually very meaningful and language-like [72] and might contain the historic record of modules which have evolved in previous environments and might be useful in the future if environmental conditions return to the previously seen state or for the control of gene expression.

In case when it is suspected that unknown engineers are originators of the organism, the text may be encoded using some unknown coding/language [108] so it might be a worthy idea to check Schelling point [93, 136] passwords [121] such as digits of π, prime numbers, Fibonacci numbers, etc.

In organisms with no DNA code or if only external observation of the sample is possible we may be interested in investigating the presence of UE of second type—functional modules which could not arise via the process of mutation with natural selection due to 'low-fitness moats' around such designs. A low-fitness moat must not just prevent evolution of the module from components; it also precludes its appearance as a reduction from a more complex module. Darwin himself put it as follows: 'if it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down' [27]. Such modules can happen by random chance only if the number of involved parts is very small, so a component with a significant number of diverse parts is unlikely to arise by chance alone.

Whether low-fitness moats exist in complicated domains, such as biology, is an open question we would like to see addressed. It is possible that they do not exist or are very rare. The argument is that the search space is so vastly high-dimensional (e.g. 3 billion base pairs in human DNA) that it is unlikely that there is literally no route through this 3-billion-dimensional space to any particular high-fitness point or region. There are similar arguments right now in deep learning about why stochastic gradient descent in large networks of millions of connections (i.e. dimensions) does not seem to be getting caught in local optima to the extent we might expect. It appears many of these 'local optima' are actually saddle points and not optima after all, and perhaps genome space is similar.

Consequently, we propose a challenge to the synthetic biology community to purposefully design and produce an organism with an unevolvable component, which meets Darwin's criteria for falsifying his theory. Can such a feat be accomplished? Can it be mathematically proven that a particular design is not evolvable, or at least statistically very unlikely? We believe those are important questions to be answered by genetic engineers and which would reconfirm the falsifiability of the Theory of Evolution [86]. If UE do not exist, every design can be naturally occurring and so it is not

possible to distinguish between natural and synthetic origins, otherwise the presence of UE can be used to prove that engineering took place.

A strong connection exists between UE of type A and type B. A long meaningful text may represent a blueprint for constructing an unevolvable organ and an unevolvable biological module can be reduced to a complex and meaningful informational pattern. By analogy with AI-Completeness [120, 127] we propose the concept of Intelligence-Completeness (I-Completeness) to indicate that certain elements are not evolvable and require intelligence to be constructed. I-complete artifacts could be reduced to other representations (text, drawing, 3D model, organism, etc) without losing their distinctive origination signature from the purposeful engineering process.

What distinguishes I-Completeness from AI-Completeness is that AI-Complete systems have no restrictions on how they can be constructed, while objects with the I-Complete property, from the definition, cannot be products of an evolutionary process. Consequently, our challenge of constructing an artificial unevolvable biological organ is equivalent to the problem of proving some problem I-Complete. From this first, hypothetical, case other problems would be shown to be I-Complete via a series of reductions, which is a well-known method in the theoretical computer science community [56].

AI-Completeness was first established [96, 127] as the property of passing the turing test (TT) [109], with other problems shown to be AI-Complete via reductions from the TT. Perhaps we can rely on the same problem for proving I-Completeness, since engineering of synthetic life requires at least human level intelligence, and that is exactly what is being detected by the TT. One possibility is to take verbatim text from someone passing the TT and to encode it in an organism's DNA, with questions from the test corresponding to specifications and answers to meaningful information. The existence of area specific TTs in domains such as art and poetry [84] suggest that we can also produce a restricted TT for the domain of genetic engineering and encode any UE descriptions as answers to questions asking to describe such structures.

3.3. Forensic evidence from the code

In theory, as long as statistical properties of samples produced by a particular algorithm can be captured, another algorithm can simulate them on purpose, essentially spoofing behavior of the original algorithm [132]. In fact, the statistical model describing the samples can serve as an engineered algorithm for generating an equivalent sample distribution be it by an evolutionary process or any other type of algorithm. Engineered algorithms are capable of both simulating natural evolution and using it as a module in achieving their goals [134]. In principal, an engineered algorithm can produce any computable distribution and so can an evolutionary algorithm with infinite computational resources, making both types of algorithms universal and claims of particular origin of samples unfalsifiable given the unlimited power of either approach. Consequently, we can never have 100 percent

certainty as to the origination algorithm, only probabilistic estimates. This analysis applies only to post-factum observations of collected samples. If we have a chance to observe and analyze the sample generator at work we can be certain as to the process used.

4. Designometry—generalization of the proposed analysis

A forensic investigator studying an explosive device, a professor looking at a plagiarized programming project, an art expert examining a potential forgery and numerous other professionals find themselves in a situation where they need to infer information about the engineer/designer/author of a product/object/text in the absence of direct access to the agent, and only in possession of the agent's output. For example, depending on the domain, the process of making such inferences is called forensic analysis [10], stylometry [3], historiometrics [102] or behavioral profiling [123, 131]. Regardless of the subdomain of inquiry, the generalized process we will call *Designometry*: is to uncover a 'signature' of the originator in the artifact and from it to identify the agent responsible or to at least learn some properties, of the design process, which produced the artifact. Designometry could be widely applied to both biological and non-biological artifacts, which are products of intentional construction. The field includes such subdomains as:

- **Artimetrics**—which identifies software and robots based on their outputs or behavior [117, 118].
- **Behavioral Biometrics**—which quantify behavioral traits exhibited by users and use resulting feature profiles to verify identity [129]. Examples of analyzed artifacts may include text, art as well as records of direct or indirect human-computer-interaction [130].
- **CAPTCHA** (**Completely Automated Public Turing test to tell Computers and Humans Apart**)—obtains input from an agent and classifies producing agent as human or artificial [26, 119].

Desingometry could itself be seen as a sub-branch of intellectology, a field proposed to 'study and classify the design space of intelligent agents, work on establishing limits to intelligence (minimum sufficient for general intelligence and maximum subject to physical limits), contribute to consistent measurement of intelligence across intelligent agents, look at recursive self-improving systems, design new intelligences (making AI a sub-field of intellectology) and evaluate capacity for understanding higher level intelligences by lower level ones' [122, 128].

Next, we will give one example, which would fall under the heading of designometry. Stylometry of text relies on statistical analysis of 'vocabulary richness, length of sentence, use of function words, layout of paragraphs, and key words' [67] to determine the gender, age [43], native language [18], personality type [70] and even intelligence of a human author or comparable properties of an artificially intelligent text

generator [4]. In general, it seems it is possible to estimate the scientific knowledge and minimum intelligence necessary to produce, or at least duplicate, a particular artifact by analyzing its complexity, prerequisite components and evidence of tools used in the production, be it an artifact/data or an abstract algorithm [58, 124]. This does not imply that anyone with the required level of intelligence would be able to produce the artifact under consideration, just that someone below that level would fail to do so. The Reader is encouraged to read about fascinating designometric analysis of the Antikythera mechanism [37], the Egyptian pyramids [52] or the Stuxnet Virus [60] for some famous examples of such efforts.

As for the anticipated future applications of designometry, one example could be given from the domain of AI Safety. Yampolskiy writes about an artificial superintelligent system confined to a restricted environment [7], which attempts to learn the nature of its designers and programmers by inspecting its own source code: '… the AI will have access to covert sources of information such as its own hardware and software and could analyze its design and source code to infer information about the designers. For example analysis of the source code may reveal to the AI that human programmers are slow (based on the file modification dates), inefficient (based on code redundancy), illogical (based on bugs in the code), have bad memory (based on the long and descriptive variable names), and do not think in code (based on unnecessary comments in the code)' [125]. Another interesting application of designometry would be to the problem of determining if the environment in which an agent (human or artificial intelligence) finds itself is natural or engineered. This has important applications in the domains of AI Safety [22], self-locating beliefs [17], life choices [46] and general philosophy [15]. Such capacity would be particularly timely as our ability to create realistic virtual worlds is improving exponentially [114]. Finally, we foresee great utilization in domain of steganography detection [103] and general forensic analysis.

Open problems in designometry include consolidation of analysis methods from specific domains, as well as development of generalized tools and tests to be used in novel domains of investigation. Man-made [44, 59], alien-made [24] artificial object detection, and exhaustive understanding of types of information which could be inferred about the originator from the artifact are all current examples of research directions in designometry. It may be useful to be able to tell if two designs were engineered by the same agent or if an agent reused parts from another design. It is also highly likely that this process could be automated via machine learning as has been demonstrated by recent work in software designometry [20].

5. Most life in the universe has engineered origins

Inspired by Bostrom's statistical argument for our universe being an engineered one [15] we suggest a similar argument in the realm of biology. Estimating the distribution in the universe of synthetic life versus naturally occurring life, it is likely that designed life (biological robots of any complexity

produced by early alien civilizations) is the significantly more common default case. Others have made similar observations, for example Dick: '…cultural evolution may have resulted in a postbiological universe in which machines are the predominant intelligence…' Dick goes on to say '… this means that we are in the minority; the universe over the billions of years that intelligence has had to develop will not be a biological universe, but a postbiological universe' [30], or Schneider specifically on high intelligence agents: '… it may be that [biologically inspired superintelligent aliens] are the most common form of alien superintelligence out there.' [95]. Similarly, Makukov *et al* state: '… at the current age of the galaxy it might be even more probable for an intelligent being to find itself on a planet where life resulted from directed panspermia rather than on a planet where local abiogenesis took place, and the Earth is not an exception from that. This is not to say that the view that terrestrial life originated locally is flawed. But subscribing largely to this view and dismissing the possibility that terrestrial life might not be a first independent generation in the galaxy is probably nothing but a manifestation of geo-anthropo-centrism (inappropriately armed with Occam's razor).' [71].

Unless evidence to the contrary exists, a given life form is statistically more likely to have its origins as a product of engineering and so our priors should be adjusted accordingly. This type of reasoning also applies to Earth: we are also likely to have our origins as synthetic life, as suggested by the theory of directed panspermia [25], seeding [74] or some other similar variants [6, 41]. In fact the approximate probability of being produced by unaided laws of physics rather than engineering is equal to one divided by the total number of self-reproducing biological robot species all the generations of intelligent beings around the universe have ever produced. In our estimate (based on Drake's equation [32]) this tends to zero as the age of the universe increases. In general, as the universe ages, the chance of any life form being an original evolved form rather than second or later generation design approaches zero. It is important to note that our statistical argument applies only to the origins of life, not to the process of speciation, which is well explained by the Theory of Evolution. In contrast, the Theory of Evolution does not make any claims regarding the origins of life.

Assuming that in our future we will seed thousands if not millions of such robot colonies (which in turn may do the same) in our quest to colonize the galaxy, we can observe that the common problem of attribution of the origins of life would show up on many planets (this would also happen under the many-worlds interpretation of quantum mechanics and as a result of robots being developed by space aliens). We may refer to this situation as the many Darwins problem (a 'Darwin' per seeded planet).

Further, let us consider a thought experiment; we shall call it the robot planet problem[3]. Suppose at some point in our future we design a very advanced humanoid (biological) self-replicating robot with the goal of exploring distant planets.

---

[3] We are aware of the Futurama episode 'A Clockwork Origin' (Episode 6, Season 9) with a similar plot.

We send a group of such robots on a long-term mission to a star known to be orbited by a number of Earth-like planets [21]. Our goal may be to establish a permanent base on one or more such planets to reserve its resources for us, in case competing alien species may have interest in the same solar system. We would also like to make the said planets habitable for human beings and to instruct our robots to await contact from their human masters. The robots are, of course, designed to be adaptable to variations in their future environment and have a general level of intelligence comparable to that of humans.

Although it may be possible to make them super-intelligent [95, 116, 122, 125], but it is probably not a rational thing to do as such robot may present a danger to us and would be harder to control [126]. Also, providing robots with very specific goals may produce undesirable side-effects and may not work well in a large number of planets with unknown conditions. Perhaps our instructions to them will be some-thing like: 'reproduce to a number sufficient to obtain full control of your host planet, make it habitable for yourself and your masters and await arrival of your designers'. A number of less important instructions can be provided, such as: maintain good condition of each robot, establish a rule of law, do not destroy other robots, etc. It is possible that the planets in question may already contain some forms of life, but probably not highly intelligent life, so additional instructions may be provided to preserve local biodiversity.

As a significant amount of time passes on the robot planet, the group's mission is probably going to progress fairly well with the construction of necessary infrastructure, increase in population and development of sophisticated local culture and religious tradition centered around its human masters. At some point, most or all robots would have no direct knowledge of their human masters. Considerable advances are likely to have been made in terms of science and technology. At this point it is likely that a 'robot Darwin' would appear, who would criticize the idea of human masters as an irrational belief and propose a naturalistic explanation for the inhabitants of the robot planet not too different from the theory of evolution. Since the robots were designed with ability to adapt to their new environment sufficient evidence for evolution would be found and it would quickly become a dominant and very reasonable explanation for the origins of the robot colony, in light of ideas presented in this paper.

6. Conclusions

In this paper, we have suggested a design for an experiment in which engineered life is as likely as natural life by normal-izing priors. The experiment is intended to test the current assumption that it is possible to determine if a given sample is produced by natural evolution while also allowing us to investigate the detectability of genetically modified and fully synthetic life forms, which are quickly becoming common due to the latest advances in genetic engineering. With

thought experiments we attempted to show that most current life is statistically more likely to have synthetic origins and shown how such theory could be tested by translating the problem to the domain of computer science. All investigated theories have fully naturalistic explanations and are com-pletely falsifiable.

In the theoretical case of unlimited resources (mostly time [110], but also multiverses) it is not possible to tell which type of algorithm is responsible for producing the collected samples, as all investigated algorithms are universal in a sense that they can eventually produce any pattern. The suggested analysis is also broadly applicable to biological and non-biological samples, essentially everything we can repre-sent as a binary string.

Developments in synthetic biology and evolutionary robotics raise a number of ethical, biosafety and security issues. In addition to potential development of novel deadly pathogens [55], genetically modified humans [100] and other organisms, we are also facing a potential runaway evolu-tionary process. An outcome of such process could be the appearance of dangerous and potentially superintelligent robots [79], which may cause human extinction in the same way that a large number of previously existing species went extinct because of the appearance of an intellectually superior species—Homo Sapiens.

We have reviewed a number of cases in which it is possible, as a result of forensic analysis, to conclusively state that a collected sample has been engineered rather than occurred naturally. Such telltale signs include: complexity in the absence of probabilistic resources, watermarking, multi-level encoding [85], support for future features, physical computation [12], evidence of degradation from the original design, and the engineer's signature, etc. It may even be possible for intelligent agents to perform this analysis on themselves to discover their origins. Synthetic life forms which may be discovered in the wild will be interesting to study, because they can have a number of features not found in naturally occurring ones, such as: backdoor control mechanisms, hidden capabilities, previously unseen features, etc. Studying designed systems may also leak information about the engineers behind the design. Methods to do so are of interest to forensic investigators, SETI scientists, stylo-metry practitioners and exobiologists. Finally, it is very important to note that confirmed detection of synthetic life, even in the wild, would not prove any non-naturalistic notions be it god(s), creationist myths or religion, only that engi-neering took place.

Acknowledgments

References

[1] Adleman L M 1979 *Time, space and randomness* Technical Report Massachusetts Institute of Technology, Laboratory for Computer Science

[2] Ahmed F E 2002 Detection of genetically modified organisms in foods *Trends Biotechnol.* **20** 215–23

[3] Ali N, Hindi M and Yampolskiy R V 2011 Evaluation of authorship attribution software on a chat bot corpus *23rd Int. Symp. on Information, Communication and Automation Technologies (ICAT2011) (Sarajevo, Bosnia and Herzegovina, October 27–29)* pp 1–6

[4] Ali N, Schaeffer D and Yampolskiy R V 2012 Linguistic profiling and behavioral drift in chat bots *Midwest Artificial Intelligence and Cognitive Science Conf.* pp 27

[5] Allton J, Bagby J and Stabekis P 1998 Lessons learned during Apollo lunar sample quanrantine and sample curation *Adv. Space Res.* **22** 373–82

[6] Arrhenius S 1903 Die verbreitung des lebens im weltenraum *Umschau* **7** 481–5

[7] Babcock J, Kramar J and Yampolskiy R 2016 The AGI containment problem *The Ninth Conf. on Artificial General Intelligence (AGI2015) (NYC, USA, July 16–19)*

[8] Back T, Fogel D B and Michalewicz Z 1997 *Handbook of Evolutionary Computation* (London: CRC Press)

[9] Beck M and Yampolskiy R 2012 DNA as a medium for hiding data *BMC Bioinform.* **13** A23

[10] Beck M B, Rouchka E C and Yampolskiy R V 2013 Finding data in DNA: computer forensic investigations of living organisms *Digital Forensics and Cyber Crime* (Berlin: Springer) pp 204–19

[11] Bennett C H 1988 Logical depth and physical complexity *The Universal Turing Machine: a Half Century Survey* (Oxford: Oxford University Press) pp 227–57

[12] Binder P and Ellis G 2016 Nature, computation and complexity *Phys. Scr.* **91** 064004

[13] Blumer A, Ehrenfeucht A, Haussler D and Warmuth M K 1990 Occam's razor *Readings in Machine Learning* (Burlington, MA: Morgan Kaufmann Publishers) pp 201–4

[14] Bornholt J, Lopez R, Carmean D M, Ceze L, Seelig G and Strauss K 2016 A DNA-based archival storage system *Proc. of the Twenty-First Int. Conf. on Architectural Support for Programming Languages and Operating Systems, ACM* pp 637–49

[15] Bostrom N 2003 Are we living in a computer simulation? *Phil. Q.* **53** 243–55

[16] Bostrom N 2004 The future of human evolution *Death and Anti-Death: Two Hundred Years After Kant, Fifty Years After Turing* (Palo Alto: Ria University Press) pp 339–71

[17] Bostrom N 2003 The mysteries of self-locating belief and anthropic reasoning *Harv. Rev. Phil.* **11** 59–73

[18] Brooke J and Hirst G 2013 Native language detection with 'cheap' learner corpora *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead: Proc. of the First Learner Corpus Research Conf. (LCR 2011) (Presses universitaires de Louvain)* pp 37

[19] Burrows M and Sutton G 2013 Interacting gears synchronize propulsive leg movements in a jumping insect *Science* **341** 1254–6

[20] Caliskan-Islam A, Yamaguchi F, Dauber E, Harang R, Rieck K, Greenstadt R and Narayanan A 2015 When coding style survives compilation: de-anonymizing programmers from executable binaries arXiv preprint arXiv:1512.08546

[21] Cash W 2006 Detection of Earth-like planets around nearby stars using a petal-shaped occulter *Nature* **442** 51–3

[22] Chalmers D 2010 The singularity: a philosophical analysis *J. Consciousness Stud.* **17** 7–65

[23] Church G M, Gao Y and Kosuri S 2012 Next-generation digital information storage in DNA *Science* **337** 1628–1628

[24] Cirkovic M M 2006 Macro-engineering in the galactic context *Macro-Engineering* **54** 281–300

[25] Crick F H and Orgel L E 1973 Directed panspermia *Icarus* **19** 341–6

[26] D'souza D, Polina P C and Yampolskiy R V 2012 Avatar CAPTCHA: telling computers and humans apart via face classification *IEEE Int. Conf. on Electro/Information Technology (EIT2012) (Indianapolis, IN, USA, May 6–8)*

[27] Darwin C 1859 *On the origin of species by means of natural selection, or*, The Preservation of Favoured Races in the Struggle for Life, London/Die Entstehung der Arten durch natürliche Zuchtwahl, Leipzig oJ

[28] Dawkins R 1986 *The Blind Watchmaker: Why The Evidence of Evolution Reveals a Universe Without Design* (New York: WW Norton & Company)

[29] De Simone A, Guth A H, Linde A, Noorbala M, Salem M P and Vilenkin A 2010 Boltzmann brains and the scale-factor cutoff measure of the multiverse *Phys. Rev.* D **82** 063520

[30] Dick S J 2003 Cultural evolution, the postbiological universe and SETI *Int. J. Astrobiol.* **2** 65–74

[31] Dong W, Yang L, Shen K, Kim B, Kleter G A, Marvin H J, Guo R, Liang W and Zhang D 2008 GMDD: a database of GMO detection methods *BMC Bioinform.* **9** 260

[32] Drake R 1972 A general mathematical survey of the coagulation equation *Top. Curr. Aerosol Res. (Part 2)* **3** 201–376

[33] Elenis D S, Kalogianni D P, Glynou K, Ioannou P C and Christopoulos T K 2008 Advances in molecular techniques for the detection and quantification of genetically modified organisms *Anal. Bioanal. Chem.* **392** 347–54

[34] Elliott J R 2010 Detecting the signature of intelligent life *Acta Astronaut.* **67** 1419–26

[35] Endy D and Zoloth L 2016 *Should We Synthesize A Human Genome?*, Available at: (https://dspace.mit.edu/handle/1721.1/102449)

[36] Everett Iii H 1957 'Relative state' formulation of quantum mechanics *Rev. Mod. Phys.* **29** 454

[37] Freeth T, Bitsakis Y, Moussas X, Seiradakis J, Tselikas A, Mangou H, Zafeiropoulou M, Hadland R, Bate D and Ramsey A 2006 Decoding the ancient Greek astronomical calculator known as the Antikythera Mechanism *Nature* **444** 587–91

[38] Garfinkel M S, Endy D, Epstein G L and Friedman R M 2007 Synthetic genomics: options for governance *Ind. Biotechnol.* **3** 333–65

[39] Gibson D G, Benders G A, Andrews-Pfannkoch C, Denisova E A, Baden-Tillson H, Zaveri J, Stockwell T B, Brownley A, Thomas D W and Algire M A 2008 Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome *Science* **319** 1215–20

[40] Gibson D G, Glass J I, Lartigue C, Noskov V N, Chuang R-Y, Algire M A, Benders G A, Montague M G, Ma L and Moodie M M 2010 Creation of a bacterial cell controlled by a chemically synthesized genome *Science* **329** 52–6

[41] Gold T 1960 Cosmic garbage *Air Force and Space Digest* pp 65

[42] Goldberg D E 1989 *Genetic Algorithms in Search, Optimization and Machine Learning* (Reading, MA: Addison-Wesley)

[43] Goswami S, Sarkar S and Rustagi M 2009 Stylometric analysis of bloggers' age and gender *3rd AAAI Conf. on Weblogs and Social Media*

[44] Gruen A, Baltsavias E and Henricsson O 2012 *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)* (Basel: Birkhäuser)

[45] Gurevich Y and Passmore G O 2012 Impugning randomness, convincingly *Stud. Log.* **100** 193–222

[46] Hanson R 2001 How to live in a simulation *J. Evol. Technol.* **7**

[47] Heider D and Barnekow A 2007 DNA-based watermarks using the DNA-Crypt algorithm *BMC Bioinform.* **8** 1

[48] Hodgson G M and Knudsen T 2006 Why we need a generalized Darwinism, and why generalized Darwinism is not enough *J. Econ. Behav. Organ.* **61** 1–19

[49] Holman C M 2007 The impact of human gene patents on innovation and access: a survey of human gene patent litigation *UMKC Law Rev.* **76** 295

[50] Hood L and Galas D 2003 The digital code of DNA *Nature* **421** 444–8

[51] Hutchison C A, Chuang R-Y, Noskov V N, Assad-Garcia N, Deerinck T J, Ellisman M H, Gill J, Kannan K, Karas B J and Ma L 2016 Design and synthesis of a minimal bacterial genome *Science* **351** aad6253

[52] Isler M 2001 *Sticks, Stones, and Shadows: Building the Egyptian Pyramids* (Norman, OK: University of Oklahoma Press)

[53] Jackson D A, Symons R H and Berg P 1972 Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli Proc. Natl Acad. Sci.* **69** 2904–9

[54] Jeschek M, Reuter R, Heinisch T, Trindler C, Klehr J, Panke S and Ward T R 2016 Directed evolution of artificial metalloenzymes for *in vivo* metathesis *Nature* **537** 661

[55] Kaiser J 2014 US halts two dozen risky virus studies *Science* **346** 404–404

[56] Karp R M 1972 Reducibility among combinatorial problems ed R E Miller and J W Thatcher *Complexity of Computer Computations* (New York: Plenum) pp 85–103

[57] Katzenbeisser S and Petitcolas F 2000 *Information Hiding Techniques for Steganography and Digital Watermarking* (London: Artech House)

[58] Koppel M 1987 Complexity, depth, and sophistication *Complex Syst.* **1** 1087–91

[59] Kumar S and Hebert M 2003 Man-made structure detection in natural images using a causal multiscale random field, computer vision and pattern recognition *Proc 2003 IEEE Computer Society Conf. on, IEEE* vol 1, I-119–I-126 2003

[60] Langner R 2011 Stuxnet: dissecting a cyberwarfare weapon *IEEE Secur. Priv.* **9** 49–51

[61] Langton C G 1997 *Artificial Life: An Overview* (Boston, MA: Mit Press)

[62] Legg S and Hutter M 2007 Universal intelligence: a definition of machine intelligence *Minds Mach.* **17** 391–444

[63] Lehman J and Stanley K O 2011 Abandoning objectives: evolution through the search for novelty alone *Evolutionary Comput.* **19** 189–223

[64] Lemarchand G A and Lomberg J 2009 Universal cognitive maps and the search for intelligent life in the universe *Leonardo* **42** 396–402

[65] Levin L A 1977 Invariant properties of informational bulks *Mathematical Foundations of Computer Science 1977* (Berlin: Springer) pp 359–64

[66] Levin L A 1984 Randomness conservation inequalities; information and independence in mathematical theories *Inf. Control* **61** 15–37

[67] Li J, Zheng R and Chen H 2006 From fingerprint to writeprint *Commun. ACM* **49** 76–82

[68] Lipson H and Pollack J B 2000 Automatic design and manufacture of robotic lifeforms *Nature* **406** 974–8

[69] Liu C C and Schultz P G 2010 Adding new chemistries to the genetic code *Annu. Rev. Biochem.* **79** 413–44

[70] Luyckx K and Daelemans W 2008 Using syntactic features to predict author personality from text *Proc. of Digital Humanities* **2008** 146-9

[71] Makukov M A 2014 Space ethics to test directed panspermia *Life Sci. Space Res.* **3** 10–7

[72] Mantegna R N, Buldyrev S V, Goldberger A L, Havlin S, Peng C-K, Simons M and Stanley H E 1994 Linguistic features of noncoding DNA sequences *Phys. Rev. Lett.* **73** 3169

[73] Mayr E and Provine W B 1998 *The Evolutionary Synthesis: Perspectives on the Unification of Biology* (Cambridge, MA: Harvard University Press)

[74] Miletić T 2015 Extraterrestrial artificial intelligences and humanity's cosmic future: Answering the Fermi paradox through the construction of a Bracewell-Von Neumann AGI *J. Evol. Technol.* **25** 56–73

[75] Minsky M L 1994 *Will robots inherit the Earth?*, Available at: (http://web.media.mit.edu/~minsky/papers/sciam. inherit.txt)

[76] Montanez G 2016 Detecting intelligence: the turing test and other design detection methodologies *8th Int. Conf. on Agents and Artificial Intelligence (Rome, Italy, 24–26 February)*

[77] Mueller S, Jafari F and Roth D 2016 A covert authentication and security solution for GMOs *BMC Bioinform.* **17** 389

[78] Nakamura H 1986 SV40 DNA—a message from ε Eri? *Acta Astronaut.* **13** 573–8

[79] Nijholt A 2011 No grice: computers that lie, deceive and conceal

[80] Nolfi S and Floreano D 2000 *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines* (Cambridge, MA: MIT Press)

[81] Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano M G and Moosburner M 2016 Design, synthesis, and testing toward a 57-codon genome *Science* **353** 819–22

[82] Packer M S and Liu D R 2015 Methods for the directed evolution of proteins *Nat. Rev. Genet.* **16** 379–94

[83] Park S-J, Gazzola M, Park K S, Park S, Di Santo V, Blevins E L, Lind J U, Campbell P H, Dauth S and Capulli A K 2016 Phototactic guidance of a tissue-engineered soft-robotic ray *Science* **353** 158–62

[84] Pease A and Colton S 2011 On impact and evaluation in computational creativity: a discussion of the turing test and an alternative proposal *Proc. of the AISB symposium on AI and Philosophy*

[85] Perez J-C 2015 Deciphering hidden DNA meta-codes-the great unification & master code of biology *J. Glycomics Lipidomics* **5** 1000131

[86] Popper K 2014 *Conjectures and Refutations: The Growth of Scientific Knowledge* (London: Routledge)

[87] Qin S, Yin H, Yang C, Dou Y, Liu Z, Zhang P, Yu H, Huang Y, Feng J and Hao J 2016 A magnetic protein biocompass *Nat. Mater.* **15** 217–26

[88] Querci M, Van Den Bulcke M, Žel J, Van Den Eede G and Broll H 2010 New approaches in GMO detection *Anal. Bioanal. Chem.* **396** 1991–2002

[89] Rathmanner S and Hutter M 2011 A philosophical treatise of universal induction *Entropy* **13** 1076–136

[90] Robinson R 2006 Mutations change the boolean logic of gene regulation *PLoS Biol.* **4** e64

[91] Ronald E M, Sipper M and Capcarrère M S 1999 Testing for emergence in artificial life *European Conf. on Artificial Life* (Springer) pp 13–20

[92] Sahai A and Waters B 2014 How to use indistinguishability obfuscation: deniable encryption, and more *Proc. of the 46th Annual ACM Symp. on Theory of Computing, ACM* pp 475–84

[93] Schelling T C 1980 *The Strategy of Conflict* (Cambridge, MA: Harvard University Press)

[94] Schmidhuber J 1997 A computer scientist's view of life, the Universe, and everything *Foundations of Computer Science* (Berlin: Springer) pp 201–8

[95] Schneider S 2016 Alien minds *Science Fiction and Philosophy: From Time Travel to Superintelligence* p 225

[96] Shahaf D and Amir E 2007 Towards a theory of AI completeness *8th Int. Symp. on Logical Formalizations of Commonsense Reasoning (Commonsense 2007) (California, March 26–28)*

[97] Shapiro J A 1993 Natural genetic engineering in evolution *Transposable Elements and Evolution* (Berlin: Springer) pp 325–47

[98] Shcherbak V I and Makukov M A 2013 The 'Wow! signal' of the terrestrial genetic code *Icarus* **224** 228–42

[99] Shin J H, Tam B K, Brau R R, Lang M J, Mahadevan L and Matsudaira P 2007 Force of an actin spring *Biophys. J.* **92** 3729–33

[100] Shulman C and Bostrom N 2014 Embryo selection for cognitive enhancement: curiosity or game-changer? *Glob. Policy* **5** 85–92

[101] Simmons D 2008 Genetic inequality: human genetic engineering *Nat. Educ.* **1** 173

[102] Simonton D K 2016 *Reverse Engineering Genius: Historiometric Studies of Superlative Talent* (New York: Annals of the New York Academy of Sciences)

[103] Sleiman M D, Lauf A P and Yampolskiy R 2015 Bitcoin message: data insertion on a proof-of-work cryptocurrency system *2015 Int. Conf. on Cyberworlds (CW), IEEE* pp 332–6

[104] Solomonoff R J 1964 A formal theory of inductive inference. Part I *Inf. Control* **7** 1–22

[105] Solomonoff R J 1964 A formal theory of inductive inference. Part II *Inf. Control* **7** 224–54

[106] Soros L and Stanley K O 2014 Identifying necessary conditions for open-ended evolution through the artificial life world of chromaria *ALIFE 14: The 14th Conf. on the Synthesis and Simulation of Living Systems* pp 793–800

[107] Suzuki T, Asami M, Hoffmann M, Lu X, Gužvić M, Klein C A and Perry A C 2016 Mice produced by mitotic reprogramming of sperm injected into haploid parthenogenotes *Nat. Commun.* **7** 12676

[108] Tsonis A A, Elsner J B and Tsonis P A 1997 Is DNA a language? *J. Theor. Biol.* **184** 25–9

[109] Turing A 1950 Computing machinery and intelligence *Mind* **59** 433–60

[110] Valiant L G 2009 Evolvability *J. ACM* **56** 3

[111] Wang L, Brock A, Herberich B and Schultz P G 2001 Expanding the genetic code of *Escherichia coli Science* **292** 498–500

[112] Watson R A and Szathmáry E 2016 *How Can Evolution Learn? Trends Ecol. Evol.* **31** 147–57

[113] Wehner M, Truby R L, Fitzgerald D J, Mosadegh B, Whitesides G M, Lewis J A and Wood R J 2016 An integrated design and fabrication strategy for entirely soft, autonomous robots *Nature* **536** 451–5

[114] Whitworth B 2008 The physical world as a virtual reality arXiv preprint arXiv:0801.0337

[115] Yachie N, Sekiyama K, Sugahara J, Ohashi Y and Tomita M 2007 Alignment-based approach for durable data storage into living organisms *Biotechnol. Prog.* **23** 501–5

[116] Yampolskiy R 2012 Leakproofing the singularity artificial intelligence confinement problem *J. Consciousness Stud.* **19** 1–2

[117] Yampolskiy R, Cho G, Rosenthal R and Gavrilova M 2012 Experiments in artimetrics: avatar face recognition *Trans. Comput. Sci.* **XVI** 77–94

[118] Yampolskiy R and Gavrilova M 2012 Artimetrics: biometrics for artificial entities *IEEE Robot. Autom. Mag.* **19** 48–58

[119] Yampolskiy R V 2011 AI-complete CAPTCHAs as zero knowledge proofs of access to an artificially intelligent system *ISRN Artif. Intell.* **2012** 271878

[120] Yampolskiy R V 2012 AI-complete, AI-hard, or AI-easy— classification of problems in AI *The 23rd Midwest Artificial Intelligence and Cognitive Science Conf. (Cincinnati, OH, USA., April 21–22)*

[121] Yampolskiy R V 2006 Analyzing user password selection behavior for reduction of password space *The IEEE Int. Carnahan Conf. on Security Technology (ICCST06) (Lexington, Kentucky, October 17–19)* pp 109–15

[122] Yampolskiy R V 2015 *Artificial Superintelligence: a Futuristic Approach* (London: Chapman and Hall)

[123] Yampolskiy R V 2008 Behavioral modeling: an overview *Am. J. Appl. Sci.* **5** 496–503

[124] Yampolskiy R V 2013 Efficiency theory: a unifying theory for information, computation and intelligence *J. Discrete Math. Sci. Cryptogr.* **16** 259–77

[125] Yampolskiy R V 2012 Leakproofing singularity—artificial intelligence confinement problem *J. Consciousness Stud.* **19** 194–214

[126] Yampolskiy R V 2016 Taxonomy of pathways to dangerous AI *30th AAAI Conf. on Artificial Intelligence (AAAI-2016). 2nd Int. Workshop on AI, Ethics and Society (AIEthicsSociety2016) (Phoenix, Arizona, USA, February 12–13th)*

[127] Yampolskiy R V 2013 Turing test as a defining feature of AI-completeness *Artificial Intelligence, Evolutionary Computing and Metaheuristics* (Berlin: Springer) pp 3–17

[128] Yampolskiy R V 2014 The universe of minds arXiv preprint arXiv:1410.0369

[129] Yampolskiy R V and Govindaraju V 2008 Behavioral Biometrics: a survey and classification *Int. J. Biometric* **1** 81–113

[130] Yampolskiy R V and Govindaraju V 2007 Direct and indirect human computer interaction based biometrics *J. Comput.* **2** 76–88

[131] Yampolskiy R V and Govindaraju V 2009 Taxonomy of behavioral biometrics *Behavioral Biometrics for Human Identification: Intelligent Applications* ed L Wang and X Geng (Hershey, PA: IGI Global) pp 1–43

[132] Yampolskiy R V and Govindaraju V 2006 Use of behavioral biometrics in intrusion detection and online gaming *Biometric Technology for Human Identification III. SPIE Defense and Security Symp. (Orlando, Florida)*

[133] Yampolskiy R V, Rebolledo-Mendez J D and Hindi M M 2014 *Password Protected Visual Cryptography via Cellular Automaton Rule 30, Transactions on Data Hiding and Multimedia Security IX* (Berlin: Springer) pp 57–67

[134] Yao X and Higuchi T 1999 Promises and challenges of evolvable hardware *IEEE Trans. Syst., Man Cybern.* C **29** 87–97

[135] Yokoo H and Oshima T 1979 Is bacteriophage φX174 DNA a message from an extraterrestrial intelligence? *Icarus* **38** 148–53

[136] Zuckerman I, Kraus S and Rosenschein J S 2011 Using focal point learning to improve human–machine tacit coordination *Autono. Agents Multi-Agent Syst.* **22** 289–316