Designometry – Formalization of Artifacts and Methods

Soenke Ziesche
Maldives National University
Male', Maldives
soenke.ziesche@mnu.edu.mv

Roman V. Yampolskiy
University of Louisville
Louisville, KY 40292, USA
roman.yampolskiy@louisville.edu

Abstract

Two interconnected surveys are presented, one of artifacts and one of designometry. Artifacts are objects, which have an originator and do not exist in nature. Designometry is a new field of study, which aims to identify the originators of artifacts. The space of artifacts is described and also domains, which pursue designometry, yet currently doing so without collaboration or common methodologies. On this basis, synergies as well as a generic axiom and heuristics for the quest of the creators of artifacts are introduced. While designometry has various areas of applications, the research of methods to detect originators of artificial minds, which constitute a subgroup of artifacts, can be seen as particularly relevant and, in the case of malevolent artificial minds, as contribution to AI safety.

Keywords: Designometry, artifacts, intellectology, minds, AI safety

Introduction

Yampolskiy (2016) introduced the field of designometry, which aims to detect signatures of originators within artifacts. Owing to the diversity of artifacts this type of research is currently pursued independently in different domains. Therefore, Yampolskiy (2016) proposes an overarching approach through synergies and particularly through consolidation of methods of analysis from specific domains. As he highlights there could be a particular demand in the future to determine the originators of a specific type of artifacts, which are artificial or engineered life forms or minds.

In this article we tackle this problem by presenting a survey of artifacts and a survey of designometry. In the survey of artifacts we summarize existing definitions and ontologies, followed by an innovative approach to describe the space of artifacts by allocating identification numbers to them. In the designometry survey we describe fields, which pursue designometry albeit not calling it that. Thereafter, we analyze the tools and methods of these fields and infer one abstract axiom as well as general heuristics for anyone trying to profile creators of artifacts, with a special focus on artificial minds. We also establish a link to the field of intellectology, which has been introduced by Yampolskiy (2015).

Artifacts

Definition

Artifacts have been a topic in philosophy for a long time. Aristotle distinguished between things "that exist by nature" and those existing "from other causes". For the latter group he names a bed or a coat as examples and calls them "artificial products" since it requires an art of making things.

In the meantime various other definitions for artifacts have been provided, which do not differ very much. For our purposes the one by Hilpinen (1993, 156–157) is sufficient: "An object is an artifact if and only if it has an author." In the literature instead of author also creator, originator or agent is used.

Ontology

While artifacts have be described already for a long time as the definitions above show, formal ontologies have been developed only much more recently[1]. Around the end of the twentieth century research towards formal knowledge representation systems intensified since this was required for various applications in the IT field. Borgo and Vieu (2009) present a detailed approach to extend formal ontologies for knowledge representation to include artifacts. Among various existing ontologies they propose that the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is best suited for artifacts[2].

Borgo and Vieu (2009) added to existing categories in DOLCE the category of Physical Artefact, which became a subcategory of Physical Endurant, thus a sibling of the categories Amount of Matter, Physical Object and Feature. Furthermore, they made use of the quality feature in DOLCE by assigning all physical artifacts a single individual quality named capacity that characterizes the capacities of the artifacts. This new quality enabled Borgo and Vieu (2009) to formalize a series of notions based on philosophical distinctions as well as commonsense intuitions.

There are various ways to categorize artifacts according to their qualities. As this is a wide topic of philosophical research[3], here only the most important categories are mentioned: An example for a **singular, concrete** object would be the Eiffel Tower in Paris. If the simulation of universes turns out to be feasible, then a simulated universe would be an extreme example for a singular, concrete object. As opposed to singular objects, which are unique, there are also artifacts, which have **more than one or many instances**. The instances of such artifacts have the same blueprint and the same characteristics. An example would be paper clips. A particular instance of this type of artifact would be in this example a specific paper clip. As opposed to concrete

[1] See for an overview: Franssen et al. (2013).
[2] See http://www.loa.istc.cnr.it/old/Papers/DOLCE2.1-FOL.pdf
[3] See for an overview: Hilpinen (2011).

objects, artifacts can be also **abstract**. An example would be a text or any object in a digital environment. The above categories can be divided further in subcategories. For example singular, concrete objects may be independent or dependent. An example for an **independent** object would be a chair. As opposed to this singular, concrete **dependent** objects cannot exist without a substrate. An example would be a tunnel under a mountain. This object cannot be detached from that mountain. In addition to numerous other criteria such as the method of manufacture, material properties or the intended use of the artifact it can be also distinguished if an artifact has one author or more than one author (**collectively produced artifacts**) or whether the artifact is **intended** by the author or **unintended**. For example, when a tailor makes a coat for his customer, her intention is to make a coat of a certain size and style, but she also produces scraps of cloth as by-products of her work. Moreover, there are also **biological artifacts**, which are based on DNA sequences that are not found in nature, i.e. artificial recombinant molecules.

Space of artifacts

As shown above, there have been various approaches to classify artifacts, yet there appears to be no attempt to describe the space of artifacts by allocating artifacts identification numbers[4].

Every ever-produced artifact can be assigned a unique identification number based on two parameters: the location[5] coordinates and time where and when the creator initiated the production of the artifact. (This also includes then unfinished or "work-in-progress" artifacts as they are already considered artifacts once the production has begun.) Already these two parameters allow for unique identification of artifacts, including those, which have many instances, e.g. paper clips. Even in factories where large numbers of paper clips are produced, the location and time of the beginning of their production differs if the coordinates are sufficiently fine-tuned.

For dependent artifacts the location never changes, e.g. the Eiffel Tower, while others are moved around, yet for identification the location coordinates are used where its creation began. In the same way unique identification numbers can be assigned to abstract artifacts. Obviously, there are many location-time pairs, which do not represent artifacts (i.e. every pair, where and when no creation of an artifact began).

Yampolskiy (2015) suggested that the Kolmogorov Complexity measure could be applied to identification numbers representing mind designs, for example, to determine whether shorter representations of the original number are possible. The same idea could be transferred to the identification numbers of artifacts.

Based on this approach of location-time pairs it can be inferred that the set of identification numbers of artifacts is finite.

[4] Similarly, Yampolskiy (2015) has described the space of possible mind designs.

[5] The specification of the location is not topic of this paper, but it should be as precise as possible. This means for artifacts produced by nanotechnology the location would be measured on an atomic and molecular scale.

As shown these two parameters (location and time) allow already for unique identification of artifacts, yet a third interesting parameter would be the creator of the artifact. Creators of artifacts have always a mind (which distinguishes artifacts from other objects created by nature), and Yampolskiy (2015) has shown that the space of potential minds is infinite, but countable, which enables us to use the identification number of a mind as third parameter and define the space of artifacts as follows:

Space of artifacts = {m, t, l | A mind m initiated the production of an artifact at the time t and the location l}

As we now have three parameters, also the possible subsets with the cardinal number two, i.e. creator-location and creator-time, allow for unique identification of artifacts. Similarly, there are many creator-location and creator-time pairs, which do not represent artifacts.

For various reasons and within different sciences it is of interest to uncover the time of creation, the location of creation and/or the creator of an artifact. The particular process to identify the **creator** of an artifact has been called "designometry" by Yampolskiy (2016).

Designometry survey

Definition

Yampolskiy (2016) introduced designometry as a subfield of intellectology (see Yampolskiy, 2015, and Ziesche & Yampolskiy, 2017) and describes it as the field of study, which aims "[…] 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". Yampolskiy (2016) highlights that designometry could become particularly relevant for potentially engineered life or minds. Looking at the creator-location-time triple defined above to identify artifacts, designometry is exploring ways to retrieve the first parameter of the triple, i.e. the creator.

The finite set of all artifacts can be split into the two subsets of all those artifacts whose creators can be identified and those artifacts whose creators cannot be identified. However, this subdivision depends on two more parameters, which are the types of minds (see Yampolskiy, 2015) who are pursuing the designometry and the times of their embodiment. Certain types of minds may have the capacity/intelligence to successfully identify the originators of certain artifacts after a certain time of embodiment, involving education, research etc, while this may be beyond the capacity/intelligence of other minds.

Essentially, designometry is a two-stage process, which looks successively at the following two questions:

1) Is an object an artifact?

2) If yes, is it possible to identify the mind who created this artifact?

Within various sciences the quest for the creator of artifacts is part of the research; a fact that motivated us to introduce classifications of artifacts above. The type of artifact differs depending on the particular science, and methodologies are usually not coordinated among the sciences. It is an aim of designometry to create synergies so that sciences could benefit from each other in their studies. Therefore, in what follows we survey all these fields in which relevant work is done and try to abstract away details about the field, while keeping the methods in an effort to extract generic methods of designometry. As we will see, some of these fields focus only on the second question above since by definition of their domain they deal with artifacts only, e.g. stylometry, while for other domains both questions are relevant, e.g. archaeology.

Archaeology / archaeometry

Archaeology is the classic science dealing with artifacts. Archaeology is restricted to artifacts produced by humans and deals mostly with concrete artifacts since over 99 per cent of the human development has occurred before abstract artifacts such as written texts or digital objects existed. Different phases can be distinguished: A significant amount of time and resources of an archaeological investigation is dedicated to survey areas of interest and to uncover artifacts, often through excavation. Only in the subsequent phase the artifacts are analyzed, for which methods of archaeometry are applied. Sophisticated techniques have been developed to determine the time of creation as well as the function of unknown artifacts, while the identification of a specific creator of an ancient artifact is in many cases neither possible nor pursued. An example for uncovering a signature within archaeometric research is provided by Labati at al. (2012) for the specific case of latent fingerprints on clay artifacts. To distinguish this domain from the other ones below it can be stated that archaeology and archaeometry do neither deal with biological artifacts, nor concentrate on abstract ones, which are both of interest for designometry.

Artimetrics

Yampolskiy and Gavrilova (2012) introduced the innovative field of artimetrics as an extension to the known domain of biometrics. In addition to be able to identify, classify and authenticate biological entities through sensors, which is the field of biometrics, the need arose to be able to do the same with their virtual representatives, i.e. embodied as well as virtual robots, software, and virtual reality agents. The field of artimetrics aims to identify such artificial entities based on their outputs or behavior. In the ontology above it was mentioned that artifacts could be abstract or virtual.

Astrobiology

Astrobiology is a science, which, among other things, examines whether extant or extinct life in the universe exists or existed, of which we are not aware, and if yes, through which methods it can be detected. One established yet up to now unsuccessful method is to monitor electromagnetic radiation for signs of transmissions from other life outside earth. To complement these efforts Freitas (1983) proposed the Search for Extraterrestrial Artifacts based on his "Artifact Hypothesis: A technologically advanced extraterrestrial civilisation has undertaken a long-term programme of interstellar exploration via transmission of material artifacts." He distinguishes four classes of potential artifacts: Astroengineering, self-replicating artifacts, passive artifacts and active probes. In this regard astrobiology would be an example where both issues above are relevant: First, objects have to be identified as artifact, which is potentially followed by the quest for the creator. However, Haqq-Misra and Kopparapu (2012) show that despite long lasting searches non-terrestrial artifacts would likely remain not be detected because of the vastness of the universe. Wright (2017) suggests considering "a prior indigenous technological species" in our solar system, thus earth or nearby planets would have to be scrutinized for artifacts of such species.

Behavioral biometrics

Due to the proliferation of interaction between humans and electric devices the need for various authentication methods increases. In addition to often used unique physiological characteristics of a user, another category are unique behavioral characteristics. The latter have been surveyed by Yampolskiy and Govindaraju (2008) and can be divided into behaviors that produce artifacts, e.g. texts, emails, sketches or programming codes, and those which do not produce outputs, e.g. car driving style, game strategies, lip movements or mouse dynamics. These artifacts in the first category may include unique signatures of the user, which can be further analyzed through stylometric methods, which are introduced below.

Forensics science

Forensic science concerns the collection and analysis of evidence linked to a criminal investigation. A subset of the evidence could be artifacts, for example an exploded or unexploded device, in which case the identification of the creator may advance the criminal investigation.

The subfield digital forensics addresses evidence found in digital devices, which becomes more and more relevant. All such evidence are abstract artifacts. Methods to identify the authors of such digital content are described below under "Code stylometry".

Stylometry

Stylometry is a domain, which initially focused on the particular artifact of written texts, i.e. the identification of an author of an anonymous or disputed text. There it is sometimes called authorship attribution. This field has progressed recently a lot due to the availability of a large digital text corpus, which can be utilized for statistical analysis. Moreover, stylometry has been extended to other areas of creative artifacts. Backer and van Kranenburg (2006) deliver an approach for the subfield of musical stylometry based on statistical pattern recognition. Various approaches also tackle the subfield of visual stylometry, for example Hughes et al. (2012), Qi et al. (2013) or Jacobsen and Nielsen (2013). The prime motive is often to find methods to authenticate songs and paintings respectively and at the same time to uncover forgery.

Code stylometry

Code stylometry is the subfield, which aims to finds methods to de-anonymise the creator of programming codes, which is also motivated to improve detection of plagiarism. Code stylometry is particularly interesting since automated methods through machine learning have been developed which can be applied to a large code corpus. Caliskan-Islam et al. (2015a) examine machine-learning methods to identify source code authors of C/C++ using coding style. A distinction has to be made whether source code or merely compiled binary code is available for analysis. Binary code is much harder to de-anonymise, which was tackled, for example, by Rosenblum et al. (2011) and further expanded by Caliskan-Islam et al. (2015b).

Simulation detection

Bostrom (2003) analyzed the likelihood that we are living in a computer simulation. If this were true, our universe itself would be an artifact (and the set of naturally occurring things in our universe would be empty) and the identification of its creator/simulator would be of high scientific interest, yet extremely difficult if not impossible. Due to these challenges there are not many (scientific) attempts in this regard. For example, Hsu and Zee (2006) argue that cosmic microwave background could be used as a communication channel with a potential creator/simulator. Beane et al. (2012) take a different approach and aim to show that a simulator could be in principle detected because of the finiteness of resources. Both these methods of designometry would differ from all the other methods listed here.

Synthetic biology

Synthetic biology is an interdisciplinary field, which is still in early stages and examines the design and construction of new biological entities or the redesign of existing biological entities (see e.g. Hutchison et. al., 2016). For various reasons it can be of interest based on the resulting biological artifact to identify its creator. The creator may have placed, for example, intentionally a signature in the DNA, which is called steganography[6] and is the practice of concealing information within unsuspicious cover carriers. Beck et al. (2013) provide an approach to find such messages, thus possibly identify the author. If the creator has left no such message, potentially methods from code stylometry could be applied, i.e. looking for specific patterns in the synthetic code to reveal the author’s identity.

---

[6] See for an overview: Katzenbeisser and Petitcolas (2000).

Designometry methodologies

In this section, we aim to present generic methods for designometry, either by deriving them from the individual methods applied by the above fields, which use designometry, or by introducing original techniques. These methods constitute of one axiom and a set of heuristics. As described, designometry is a two-stage process, each stage using different approaches, which is reflected in the structure of this section.

a. Is an object an artifact?

Some of the fields above have to address this question first, e.g. archaeology, while others do not have to since by definition they exclusively deal with artifacts, e.g. stylometry. As described above, even for biological samples this question may be relevant and they may turn out as artifacts. While synthetic biology deals with the manufacture of new biological entities, the reverse question how to separate engineered from natural biological entities has not received a lot of attention in the literature of the field. Yampolskiy (2016) stresses that precisely this issue could become very important in the future because of expected advances in synthetic biology, which may, however, lead to unforeseeable and undesirable consequences.

We propose the following axiom to determine whether an object is an artifact:

Axiom 1

An object is an artifact if it contains writing.

We define writing as a means of communication that uses signs. Therefore, if written signs can be found in an object, we conclude that the object is an artifact. Signs are studied in the field of semiotics. We use here the dyadic[7] approach by Saussure (1983), which distinguishes between a signifier, i.e. the form a sign takes, and the signified, i.e. the concept a signifier refers to. Written signs can be iconic or symbolic[8]. Iconic signs are characterized by a similarity between signifier and signified. An example is a drawing of a chair. Symbolic signs are arbitrary and the relation between signifier and signified is conventional within a particular language. An example is the writing of the word "chair". This means symbolic signs can be anything as long as there is a social convention about its meaning[9].

Symbolic signs, which are the main subject of Saussure's research, create for the above axiom the challenge to identify writing in an artifact. This challenge would be particularly hard for

[7] An alternative in semiotics is the triadic approach by Peirce (1958).

[8] Another category are signs that are indexical. These are signs where the signified causes the signifier. For example, fire causes smoke; in other words, smoke signifies fire. However, this category does not apply to written signs.

[9] In semiotics it is usually argued that also the grasping of the meaning of iconic signs involves to some extent social conventions.

artifacts, which were created by a type of mind whose written language humans do not know. Using Saussure's terms, the two-fold challenge is not only to get an understanding of the signifier, but before to determine whether any material within an artifact constitutes at all a signifier.

Ways to tackle this challenges are explored in the subfield exosemiotics. Exosemiotics studies signs that theoretically could be understood by other minds (see e.g. Reed, 2000). An early attempt is the constructed language Lincos, which Freudenthal (1960) created to be understandable by any possible intelligent mind. Yet the space of possible minds is vast (see e.g. Yampolskiy, 2015) and so could be potentially their respective use of signs.

Once an object has positively been declared to be an artifact the second designometric query has to be tackled, which we approach by the following heuristics:

b. Is it possible to identify the mind who created this artifact?

Heuristic b.1

One option to identify the mind who created the artifact is to look for an <u>intentional</u> signature.

This heuristic is relevant for artifacts where the creators intentionally included writing, which allows backtracking to the creator. In trivial cases these could be literally signatures or other physical or digital watermarks, which serve for copyright protection. Yet these signatures could be also forgeries, which is another field of research. Intentional signatures on artifacts are relevant in forensic science as they may serve as claim of responsibility for a misdoing. One example would be illegal graffiti with so-called tagging as a form of personal expression. Likewise, forensic scientists have to be aware of forgeries. A special case are intentional, but hidden signatures, which was introduced above as steganography and is explored for example in synthetic biology.

In some contexts the identification of a species as creator instead of an individual is already sufficient. This is the case quite often in archaeometric research, but also in astrobiology it would be an unprecedented success if an artifact with a signature of a species not originating from earth could be found. An example for such an artifact would be the so-called Pioneer plaques, which travelled to space on board the Pioneer 10 and Pioneer 11 space probes in 1972 and 1973 respectively. These plaques contain a pictorial message, based on exosemiotic considerations, for a potential encounter with extraterrestrial life, for whom it is intended to be sufficient to identify the human species as creator of this artifact rather than the human individuals involved in the production of the plaques.

Heuristic b.2

One option to identify the mind who created the artifact is to look for an unintentional signature.

While intentional signatures often signal that the creator wants to be linked to this artifact and no scientific work for identification is required, the situation is different when the creator obviously does not want to be identified and searching for unintentional signatures is used as a method to do so.

This heuristic is even more common in forensic science than b.1 since usually culprits do not want to be identified. Therefore, methods of forensic science focus on the revelation of unintentional signatures at artifacts linked to a crime such as fingerprints or more recently DNA. The search for signatures is not a priority within archaeometric research, but one example by Labati at al. (2012) about latent fingerprints on clay artifacts was mentioned above.

Other types of unintentional signatures would be certain patterns, which could be unambiguously retraced to the creator. Two categories can be distinguished: Whether it is attempted to detect the pattern from the design and/or the behavior of the artifact or from the underlying code, which defines the artifact.

Examples for the first category would be artimetrics and behavioral biometrics. Here, a black box approach is applied and the search for patterns focuses on the observed properties, output or behavior. Another example are the nests of weaverbirds. Bailey et al. (2015) describe how nests can be attributed to individual weaverbirds through computer-aided image texture classification. The signatures have to be considered unintentional if we assume that the minds of the creators are not sufficiently sophisticated. Also in certain contexts the identification of a species through unintentional signatures instead of an individual is already scientifically satisfying. This is the case for many animal-built structures[10].

An example for the exploration of patterns in the code would be code stylometry. In the context of biological entities the distinction above could be formulated as to search for a signature in the phenotype or in the genotype of the organism. Overall, the availability of a code or genotype is preferred since this allows performing more precise analysis of a specific dataset.

If the desirable code is not at hand another tool related to designometry becomes relevant, which is **reverse engineering**. This is the process of taking an artifact and trying to obtain knowledge about its construction plan or its code. Reverse engineering can be hard and costly. Villaverde and Banga (2014) provide an overview of available reverse engineering methods for biological entities, which is relevant since our special interest lies in designometry related to engineered life.

[10] See for an overview: Hansell (2005).

Therefore, reverse engineering can be seen as a function from the space of artifacts, represented by the set of identification numbers of artifacts, to the code of the artifact, whereby the artifacts are represented by its identification number:

re: {artifacts} -> {code}

Artifacts with more than one instance have individual identification numbers as defined above, but share the same code, thus this function is not bijective.

Heuristic b.3

One option to limit the set of minds who created the artifact is to focus on those minds who have a motive, goal or drive to create this artifact (and to potentially sign it).

Artifacts are created for a reason, thus the potential creators of a particular artifact can be narrowed down to the subset of minds that have a motive, goal or drive to create this artifact. In forensic science potential motives for a crime play an important role. According to Bostrom's (2012) orthogonality thesis the goals and the intelligence levels of minds are independent of each other; hence the heuristics 7 and 8 are independent. Regarding artifacts, which are crated by artificial minds, Omohundro (2007 and 2008) defines general drives for such activities: Efficiency, self-preservation, acquisition, and creativity. This heuristic addresses also potential writing on the artifact, which is relevant for the above axiom and heuristics, as this may reveal a sub-goal by the creator to not only produce the artifact, but also to sign it.

Heuristic b.4

One option to limit the set of minds who created the artifact is to focus on those minds who have the necessary skillset, education and intelligence to create this artifact.

Apart from artifacts, which minds can create only because of their genetic predisposition, it requires intelligence as well as education to create certain artifacts, which only a subset of minds acquires. For example, for a long time only Chinese and Japanese had the knowledge and skills to produce porcelain, while the production of European porcelain only started in 18th century. Therefore, the set of creators of any porcelain produced before 1700 can be limited to Chinese and Japanese minds.

The minds with a certain skillset to create certain artifacts could be subsets within a species, but also whole species. Above we introduced animal-built structures, and for example, if we look at the artifact spider webs, it has many instances, but the set of potential creators is limited to the set of spiders.

It has to be noted that the creator does not have to be the inventor of a particular artifact, but has acquired the skills to produce the artifact. The inventor of an artifact with more than one instance could be identified by checking, which creator-location-time triple of all instances of this artifact has the lowest/earliest time number, thus to determine when this artifact was produced for the first time[11].

The degree of skills required to produce an artifact varies highly. It can be inferred that the more sophisticated the artifact is, the smaller is the subset of minds who is able to create this of artifact. As designometry aims to focus on sophisticated artifacts such as artificial minds in particular, this heuristic could be useful.

Heuristic b.5

One option to limit the set of minds who created the artifact is, if the time and the location coordinates when and where the production of the artifact was initiated can be determined, to look only at minds who were embodied at this time and were able to reach this location.

As established above, all artifacts can be defined through a creator-location-time triple, and designometry addresses the quest for the creator. However, the other two parameters could be supportive as this heuristic shows.

This heuristic is for example applied in archaeology and also in forensic science. While in archaeometric research it is often satisfying to limit the set of creators to a certain epoch, in forensic science the manhunt is for specific individuals. Therefore, the whereabouts of suspects in relation to artifacts, which are linked to the investigation, constitutes critical information.

Heuristic b.6

One option to identify the mind who created the artifact is to look for witnesses, testimonies, coverage, footage, logfiles or sensor data of the creation of the artifact.

According to this heuristic information about the creator of an artifact could be gathered if another mind or another sufficiently equipped artifact has witnessed the creation of the artifact. On this heuristic we have to rely especially if the artifact does not exist anymore. One example would be the Seven Wonders of the Ancient World, of which nowadays only the Great Pyramid of Giza still is in existence. That the other six also existed and that for example the creator of the Statue of Zeus at Olympia was a Greek sculptor named Phidias we have to believe because of testimonies.

For testimony by humans the criterion of credibility is relevant. However, more and more features and activities in the physical world are captured electronically e.g. through sensors. These data naturally include also evidence about creation of artifacts. This could be relevant for the creation of artifacts, which involves computers such as artificial minds, as logfiles or similar digital traces may link to the creator.

[11] An exceptional case are artifacts that were invented more than once independently at different times and dates. An example is the calculus, which is credited to both Newton and Leibniz.

Although the introduced axiom and heuristics mark an early phase of designometry the envisaged contributions of designometric methods can be summarized as follows:

- To provide synergies to fields, which have the overlapping sub-goal to identify creators of artifacts, but which did not cooperate much yet, and enhance as well as consolidate their efforts.
- To provide time-critical input to AI safety from an innovative angle as the heuristics are applicable to the subset of artificial minds.

Designometry as a sub-branch of intellectology

As Yampolskiy (2016) suggested designometry can be seen as a subfield of intellectology, which was introduced by Yampolskiy (2015) in order to examine in more detail features of the space of minds. More precisely, designometry can be seen as a function from the space of artifacts to the space of minds.

d: {artifacts} -> {minds}

Since we have shown that artifacts can have identification numbers assigned to and Yampolskiy (2015) has done the same for the space of minds this is a function between natural numbers:

d: ℕ -> ℕ

This means the designometry function maps the unique identification number of an artifact to the unique identification number of the mind of its creator. A special case is the category "collectively produced artifacts", which was introduced above. These artifacts have more than one creator, i.e. the identification number of the artifact is mapped on a set of numbers, which are the identification numbers of all its creators.

As shown above an intermediate stage is often (but not always) to map the artifact to its code, in which then hints for the mind, who created the artifact, may be found.

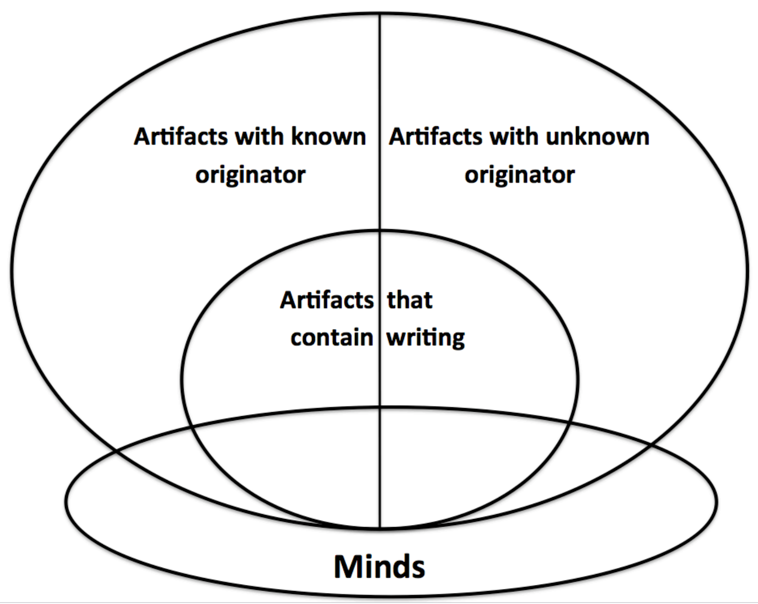
The introduction of the creator-location-time triple enables us to define further interesting sub-categories. Creator-time pairs for artifacts are only possible during the time period when the creator with this particular mind has been embodied; a fact that is utilized in heuristic b.5 above. An interesting set is, for example, the set of all artifacts produced by a particular mind during its embodiment.

Of particular relevance is the set of artifacts that comprises all artificial minds, which can be also referred to as the set of not naturally occurring minds within the space of possible minds, and the question which minds are capable to create them (see heuristic b.4 above)? In other words, for the set of artificial minds what is the set of potential creators within the creator-location-time triple?

Artificial minds constitute the intersecting set of the set of minds and the set of artifacts. Since artificial minds may also produce other artifacts, a nested constellation emerges. This raises further interesting questions: What is the set of artifacts that can only be produced by artificial minds, but not by naturally occurring minds? Are there artifacts, which can only be produced by those artificial minds, which themselves have been produced by artificial minds? This question can be applied to deeper levels of nesting too. In this regard it makes probably sense to distinguish between offsprings of artificial minds resulting from reproduction, which are by definition also artifacts, and those artifacts, which are produced by artificial minds outside of reproduction. One trivial conclusion is that the time parameter of the creating mind must be lower than the time parameter of the created mind.

Through the introduced methodologies of designometry various sets and subsets of artifacts can be defined as partly illustrated in Figure 1.

Figure 1: Categories of artifacts including artificial minds[12]



Conclusion and future work

We have presented two surveys, a survey of artifacts and a survey of designometry. We demonstrated how these surveys are interconnected. The new field of designometry aims to find general tools and methods to identify the creator of artifacts. Currently this field is divided

[12] The sizes of the areas do not represent proportions of their cardinality.

in specialized subfields for particular purposes, e.g. criminal investigation, or particular artifacts, e.g. fine art.

We aimed to provide a bigger picture by specifying the space of artifacts through a creator-location-time triple and extracted the two-stage process that first it has to be ascertained that an object is an artifact and then the quest for the creator has to be tackled. For both stages we have formulated an axiom and general heuristics, some of which were partly and rather implicitly applied before and some of which are innovative. We have determined that writing on an object is a clear indicator that the object is an artifact. For the identification of the creator the examination of the code for signs appears to be more promising than scrutinizing properties or behavior of an artifact. Therefore, procedures such as reverse engineering to obtain the code are a relevant instrument for designometry.

Yampolskiy (2016) proposed the field of designometry with an outlook towards a particular subset of artifacts, which are artificial minds. Given the ongoing progress in the concerned technologies, such research is very timely. Methods to find out the creator of artificial minds are likely to be relevant for several reasons, e.g. for proper registration, but can be seen in particular as a contribution to the field of AI safety, which entails the identification of originators of malicious systems as a critical step to curtail such systems if possible. Both the specification of the space of artifacts as well as the proposed axiom and the initial set of heuristics to identify creators of artifacts can be seen as groundwork for future AI safety research.

Regarding future work, in addition to the proposals and open questions, which were mentioned above, we may look at a fourth parameter to define the space of artifacts, supplemental to creator, time and location, which would be the goal or purpose of the particular artifact. In heuristic b.3 above we looked at the goals of the creators of artifacts, which could be connected to the goals of the artifacts, referred to by Tegmark (2017) as outsourcing of goals through engineering. The feature "intended use of the artifact" has been introduced before within the ontology of artifacts, but it had not been linked to AI safety. In addition to establishing who created artifacts and when and where, it is many contexts important to know whether the artifact has benevolent or malevolent goals.

Tegmark (2017) highlights the increasing relevance of goals of artifacts since in addition to living organisms having goals a "[…] rapidly growing fraction of matter was rearranged by living organisms to help accomplish their goals." He also presents data that show that "[…] most matter on Earth that exhibits goal-oriented properties may soon be designed rather than evolved". This interesting observation motivates us to expand our work towards goals of artifacts.

Finally, it has to be reiterated as in Yampolskiy (2016) that when we discuss engineered life and artificial minds, we do not support by any means non-naturalistic notions be it god(s), creationist myths or religion, but merely the engineering of biological entities.

References

Aristotle, Physica, in The Works of Aristotle Translated into English (Volume II), David Ross (ed.), Oxford: Clarendon Press, 1930.

Backer, E., & van Kranenburg, P. (2005). On musical stylometry—a pattern recognition approach. *Pattern Recognition Letters*, *26*(3), 299-309.

http://www.sciencedirect.com/science/article/pii/S0167865504003393

Bailey, I. E., Backes, A., Walsh, P. T., Morgan, K. V., Meddle, S. L., & Healy, S. D. (2015). Image analysis of weaverbird nests reveals signature weave textures. Royal Society open science, 2(6), 150074.

http://rsos.royalsocietypublishing.org/content/2/6/150074

Beane, S. R., Davoudi, Z., & Savage, M. J. (2014). Constraints on the Universe as a Numerical Simulation. *The European Physical Journal A*, *50*(9), 148.

https://arxiv.org/pdf/1210.1847.pdf

Beck, M. B., Rouchka, E. C., & Yampolskiy, R. V. (2012, October). Finding data in DNA: computer forensic investigations of living organisms. In *International Conference on Digital Forensics and Cyber Crime* (pp. 204-219). Springer, Berlin, Heidelberg.

https://pdfs.semanticscholar.org/c8c3/fe9bb61cf44b29a1bdc1e40cdb5a6894c978.pdf

Borgo, S., & Vieu, L. (2009). Artefacts in formal ontology. *Handbook of philosophy of technology and engineering sciences*, *9*, 273-307.

ftp://ftp.irit.fr/IRIT/LILAC/BV-HBPT09.pdf

Bostrom, N. (2003). *Are You Living In a Computer Simulation?* Philosophical Quarterly. 53(211): p. 243-255.

https://ora.ox.ac.uk/objects/uuid:44c386c4-5d9e-4ecf-a47c-9631a2a59747/datastreams/ATTACHMENT01

Bostrom, N. (2012). The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents. In V.C. Müller (Ed.), Theory and Philosophy of AI. Special issue, Minds and Machines, 22 (2), 71-85.

Caliskan-Islam, A., Harang, R., Liu, A., Narayanan, A., Voss, C., Yamaguchi, F., & Greenstadt, R. (2015, August). De-anonymizing programmers via code stylometry. In *24th USENIX Security Symposium (USENIX Security), Washington, DC*.

https://www.cs.drexel.edu/~ac993/papers/caliskan_deanonymizing.pdf

Caliskan-Islam, A., Yamaguchi, F., Dauber, E., Harang, R., Rieck, K., Greenstadt, R., & Narayanan, A. (2015b). When coding style survives compilation: De-anonymizing programmers from executable binaries. *arXiv preprint arXiv:1512.08546*.

https://arxiv.org/pdf/1512.08546.pdf

Franssen, M., Kroes, P., Reydon, T., & Vermaas, P. E. (Eds.). (2013). *Artefact kinds: Ontology and the human-made world* (Vol. 365). Springer Science & Business Media.

Freitas, R. A. (1983). The search for extraterrestrial artifacts(SETA). *British Interplanetary Society, Journal (Interstellar Studies)*, *36*, 501-506.

http://www.rfreitas.com/Astro/SETAJBISNov1983.htm

Freudenthal, H. (1960). Lincos, Design of a Language for Cosmic Intercourse. Amsterdam: North-Holland.

Hansell, M.H. (2005). Animal Architecture. Oxford University Press, Oxford pp 321.

Haqq-Misra, J., & Kopparapu, R. K. (2012). On the likelihood of non-terrestrial artifacts in the Solar System. Acta Astronautica, 72, 15–20.

https://arxiv.org/pdf/1111.1212.pdf

Hilpinen, R. (1993). Authors and artifacts. In *Proceedings of the Aristotelian Society* (Vol. 93, pp. 155-178). Aristotelian Society, Wiley.

Hilpinen, R. (2011). Artifact. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

https://plato.stanford.edu/entries/artifact/

Hsu, S., & Zee, A. (2006). Message in the Sky. *Modern Physics Letters A*, *21*(19), 1495-1500.

https://arxiv.org/pdf/physics/0510102.pdf

Hughes, J. M., Mao, D., Rockmore, D. N., Wang, Y., & Wu, Q. (2012). Empirical mode decomposition analysis for visual stylometry. *IEEE transactions on pattern analysis and machine intelligence*, *34*(11), 2147-2157.

http://ieeexplore.ieee.org/abstract/document/6127875/

Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., ... & Pelletier, J. F. (2016). Design and synthesis of a minimal bacterial genome. *Science*, *351*(6280), aad6253.

https://pdfs.semanticscholar.org/9c1c/932fca27afa2ada8c1653ce9d22500e1abe6.pdf

Jacobsen, C. R., & Nielsen, M. (2013). Stylometry of paintings using hidden Markov modelling of contourlet transforms. *Signal Processing*, *93*(3), 579-591.

http://people.math.aau.dk/~mnielsen/reprints/2013_stylometry.pdf

Katzenbeisser, S., & Petitcolas, F. (2000). Information hiding techniques for steganography and digital watermarking. Artech house.

Labati, R. D., Genovese, A., Piuri, V., & Scotti, F. (2012). Two-view contactless fingerprint acquisition systems: a case study for clay artworks. In *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2012 IEEE Workshop on* (pp. 1-8). IEEE.

http://piurilabs.di.unimi.it/Papers/bioms_2012.pdf

Omohundro, S.M. (2007). *The Nature of Self-Improving Artificial Intelligence*. Paper presented at Singularity Summit 2007, San Francisco, CA.

Omohundro, S.M. (2008). The Basic AI Drives. In P. Wang, B. Goertzel & S. Franklin (Eds.), *Proceedings of the First AGI Conference, Frontiers in Artificial Intelligence and Applications, Volume 171*. Amsterdam: IOS Press, 483-492.

Peirce, C.S. (1958): *Collected Writings* (8 Vols.). (Ed. Charles Hartshorne, Paul Weiss & Arthur W Burks). Cambridge, MA: Harvard University Press.

Qi, H., Taeb, A., & Hughes, S. M. (2013). Visual stylometry using background selection and wavelet-HMT-based Fisher information distances for attribution and dating of impressionist paintings. *Signal Processing*, *93*(3), 541-553.

http://www.sciencedirect.com/science/article/pii/S0165168412003623

Reed, M. L. (2000). Exosemiotics: an inter-disciplinary approach. *Acta Astronautica*, *46*(10), 719-723.

Rosenblum, N., Zhu, X., & Miller, B. (2011). Who wrote this code? Identifying the authors of program binaries. *Computer Security–ESORICS 2011*, 172-189.

http://ftp.cs.wisc.edu/paradyn/papers/Rosenblum11Authorship.pdf de Saussure, Ferdinand (1983): *Course in General Linguistics* (trans. Roy Harris). London: Duckworth

Tegmark, M. (2017). Life 3.0 Being Human in the Age of Artificial Intelligence. New York: Knopf.

Villaverde, A. F., & Banga, J. R. (2014). Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of the Royal Society Interface*, *11*(91), 20130505.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3869153/

Wright, J. T. (2017). Prior indigenous technological species. *International Journal of Astrobiology*, 1-5.

https://arxiv.org/pdf/1704.07263.pdf

Yampolskiy, R. V. (2015). The space of possible mind designs. In *International Conference on Artificial General Intelligence* (pp. 218-227). Springer International Publishing.

https://www.researchgate.net/profile/Roman_Yampolskiy/publication/300646188_The_Space_of_Possible_Mind_Designs/links/5737f1b408ae9ace840bfa3a.pdf

Yampolskiy, R. V. (2016). On the origin of synthetic life: attribution of output to a particular algorithm. *Physica Scripta*, *92*(1), 013002.

https://www.researchgate.net/profile/Roman_Yampolskiy/publication/310736484_On_the_origin_of_synthetic_life_Attribution_of_output_to_a_particular_algorithm/links/5836117708aec3fe331c5203/On-the-origin-of-synthetic-life-Attribution-of-output-to-a-particular-algorithm.pdf

Yampolskiy, R. V., & Gavrilova, M. L. (2012). Artimetrics: Biometrics for Artificial Entities." *IEEE Robotics & Automation Magazine* 19, no. 4, 48-58.

Yampolskiy, R. V., & Govindaraju, V. (2008). Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, *1*(1), 81-113.

https://pdfs.semanticscholar.org/1958/eebd997a2e90b88e1f8bb5345ec88408a1ce.pdf

Ziesche, S., & Yampolskiy, R. V. (2017). High Performance Computing of Possible Minds. *International Journal of Grid and High Performance Computing (IJGHPC)*, *9*(1), 37-47.