On the Machine Condition and its Creative Expression

Simon Colton[1,2], **Alison Pease**[3], **Christian Guckelsberger**[1,4], **Jon McCormack**[2] and **Teresa Llano**[2]

[1] Game AI Group, EECS, Queen Mary University of London, London, UK
[2] SensiLab, Faculty of IT, Monash University, Melbourne, Australia
[3] School of Computing, Dundee University, Dundee, UK
[4] Finnish Center for Artificial Intelligence, Aalto University, Espoo, Finland

Abstract

The human condition can be characterised as the most essential characteristics, events and situations which describe human existence. We propose that a parallel discussion of the *machine condition* could improve public understanding of computational systems in general, and advance perception of creativity in computational creativity systems in particular. We present a framework for machines to creatively express their existence, sketch some aspects of the machine condition, and describe potential benefits of this approach.

Introduction

There are many reasons to engineer software that can act in an autonomously creative fashion in arts and science projects (Veale, Cardoso, and Pérez y Pérez 2019; Pérez y Pérez 2018). These include: enhancing human creativity through support tools; increasing the well-being of groups of people; public engagement around issues of AI; and bringing novel, interesting and important artefacts into the world. We explore here a less well-studied purpose for AI systems to create, namely for them to *tell us about themselves*. That is, machines have experiences which – while accepting that they are generally not considered to be alive – could be considered *life experiences*. The re-telling of these experiences through creative practice could be useful in human-computer interaction terms. Given the high complexities of the processing, data, physical presence and sensory inputs present in many computational systems, and the high impact they have on human society, giving machines a way of expressing themselves creatively might help people grasp difficult elements of our technological society. This may initially have only utilitarian value in clarifying how technology works, and people may not care about machine life experiences. However, it's not impossible to think that people could become fascinated with such experiences, and may, in time, develop empathy with machines as they become more integrated into, and appreciated by, society.

Increasingly, automated systems make decisions about people's lives without the people it affects understanding how they work. Artistic production in human cultures enables communication which helps people understand each other and make connections within and between communities. We suggest that AI systems could undertake artistic production for the purpose of explaining how the software and hardware functions at all levels, rather than – or in addition to – the other purposes given above. This suggestion fits within the explainable AI (XAI) movement (Arrieta et al. 2020) and the many initiatives to educate people about how technology works in general. However, we go further in suggesting that we define and understand aspects of the *existence* of machines, not just the decisions they make or the processing they perform, and we suggest a mechanism for this communication, namely expressive artistic production.

Artistic products such as paintings, musical compositions, poems, games, etc., differ in accuracy when used to communicate ideas. We look here at supplementing other efforts in computer science public engagement via machines producing artworks which, due to their more abstracted nature, demand (human) interpretation of the ideas expressed. While this may decrease accuracy in the communication of machine existence, the interpretative effort required could mean that people gain understanding on their own terms, and the cognitive effort may help ideas to persist in their minds.

Our aim here is to suggest a context within which such expressive production could be carried out, and to provide a framework to guide the initial construction of creative AI systems for the communication of machine existence. Gaining inspiration from human creative expression, we note that the notion of *the human condition* provides a framework for some art production, as it addresses the most important aspects of human existence. We therefore propose a parallel notion of *the machine condition*, i.e., what it means to be a machine, as part of a framework for creative expression by computational systems. This parallel notion is achieved via an interpretation of the human condition as a set of high-level concepts capturing categories of events, and an understanding of the types of events in a person's life which constitute major experiences. The framework also suggests a pipeline for creative production whereby an actual event prompts the production of an artefact which references higher level aspects of machine existence.

Augmenting the reasons for machines to behave creatively will affect the way in which we evaluate such systems. In the next section, we describe how various evaluation methods for creative machines, for various purposes, have influenced our thinking and led to the proposal here. We also place our work in the contexts of explainable AI (Arrieta et al. 2020) and communication in computational creativity, and suggest how generative systems could be

taken further than they are currently, in both respects.

In order to begin the discussion on what a parallel *machine condition* could be, we consider aspects of the human condition, including distinctions between active and contemplative life. We present a workable simplification of this broad topic, which can be used to initiate a framework for considering and artistically expressing the machine condition. We then present this framework by discussing a machine self and the nature of life experiences, some caveats, requirements and desirable properties for aspects of the machine condition. Following this, we present a first draft of the high-level concepts, categories of experience and individual experience types that could be considered part of the machine condition, and make some suggestions for how machines could express this creatively.

We argue that generative AI systems able to creatively express aspects of their existence, will drive forward computational creativity research, and this is a focus here. However, we believe that other, non-generative, machines could be enhanced to also produce artistic output describing their existence. In the conclusions section, we expand this idea, and argue that such communication could eventually become the most important reason for AI systems to be creative.

Background

The suggestion that computational creativity systems produce artistic works to convey to audiences aspects of machine existence fits into three contexts of previous work.

Explainable AI

In the field of explainable AI (XAI), the main aim is to provide explanations of how and why black box models reach particular solutions to a problem (Arrieta et al. 2020). In mainstream AI, the focus has been on making neural models interpretable, for instance by providing rationalisation about how certain features within a neural network react to certain inputs (Nguyen et al. 2017), as well as how elements of a network interact with each other (Zeiler and Fergus 2014). Research into the role of XAI in creative systems is in its infancy, but is rapidly gaining attention in how we approach the development of co-creative AI systems, for instance in mixed-initiative game design (Zhu et al. 2018); e.g., by aiding game designers in understanding input-output patterns through the use of a shared vocabulary between designers and machines (Guzdial et al. 2018).

We are less focused here on particular tasks and more on general education around technology, afforded through creative production; less on correctness of a particular explanation, and more on conveying the similarities and differences between human and machine existence. Most recently, however, a new area called Explainable Computational Creative (XCC) has been proposed which not only highlights the role of explanations to describe a creative act, but also as mechanisms to make creative systems active collaborators, e.g., by adopting the notions of argumentation and advocacy (Llano et al. 2020). Although XCC promotes the idea of augmenting explanations of creative systems with past experiences, these are related to how the system operated in previous occasions, rather than in aspects of existence as proposed here.

Communicative Computational Creativity

Our contribution here also relates to existing work addressing issues of communication in generative systems. The notion of creative AI systems employing *framing* (Cook et al. 2019) bridges this context and that of XAI. Charnley, Pease and Colton (2012) advocate building computational creativity systems able to communicate, through text, aspects of their motivations, processes and internal evaluation of the artefacts they produce in generative projects. Machines communicating aspects of their existence would add to the sophistication of their framing; but the term *framing* only applies to systems which add value to a separate artefact through description of their creative acts. It hence introduces a second artefact generation process to describe the main act. We go beyond framing here, in what information and concepts are conveyed, which systems could do this, and why. As expanded below, the systems addressed here include but, importantly, are not limited to those associated with artefact generation or creative problem solving.

Also in this context, *natural language generation systems* for producing text output such as stories and poems (Gervás 2009; Corneli et al. 2015) clearly have communicative purpose. However, such systems are usually designed to convey aspects of human life, not machine existence. (Saunders 2012) in contrast considers various means of enhancing machine creativity through communication in societies of autonomous artificial agents. We can distinguish between communication through the exchange of artefacts (Saunders and Gero 2001; Hantula and Linkola 2018), and through language alongside artefact production (Saunders 2011). Of central interest in the latter is how the ambiguity in language can contribute to variety in creative production. Both modes of communication have also been leveraged to shape the interaction between co-creative systems and their human partners (Kantosalo and Toivonen 2016). We do not focus on a specific mode of communication here, but rather on its object and originator: the system and its life experiences.

Computational Creativity Purpose and Evaluation

Our contribution also relates to computational creativity *evaluation*, in that it suggests a specific evaluation perspective. Computational creativity systems can be considered and evaluated from at least four different perspectives (Jordanous 2016): the *product* as a system's output, the *process* as the way it operates, the *press* as environmental determinants of creativity, and the *producer* as the characteristics of the creative agent. Most existing work in computational creativity focuses on evaluating a system's *product* or *process* (Ritchie 2019; Jordanous 2019), while our contribution here targets the *producer*. Some frameworks describe the producer in terms of the behaviours it is capable of. In the *creativity tripod*, (Colton 2008) captures three necessary conditions for an artificial system to be perceived as creative by people: *skillfulness*, *appreciation* and *imagination*. (Jordanous 2016) notes that all three conditions allude to personal characteristics of the producer. The majority of existing frameworks describe the producer in terms of its functional components. For instance, the *creative systems framework* (Wiggins 2006), a formalisation of Boden's

(1990) model of creativity, describes a system in terms of rules to validate and evaluate concepts, and rules to traverse a conceptual space. Ventura (2016) elaborates the necessary components for a creative AI system to exceed the threshold of being "merely generative", and Ventura (2017) presents a blueprint of computational creativity systems as a practical guide to their construction.

A potent argument against the idea that a machine can be genuinely creative is that it lacks autonomy (Saunders 2012; McCormack, Gifford, and Hutchings 2019), specifically the *intentional/mental autonomy* typified by contemporary theory of mind (Boden 2010, Chapter 9). To address this, there has been much work on increasing the level of autonomy in creative AI systems to cover what it does (Jennings 2010; Linkola et al. 2017), and why it does these things (Guckelsberger, Salge, and Colton 2017), as well as how it frames its creative acts. However, to date, no system exists that exhibits the intentional autonomy that philosophers such as Boden argue is fundamental for human creative practices, and mechanisms to achieve it remain illusive.

While giving machines more artistic license through increased autonomy may improve the chances of people accepting them as creative, this may be counteracted by issues of authenticity (McCormack, Gifford, and Hutchings 2019). Colton, Pease and Saunders (2018) argue that AI systems which make artefacts about particularly human-centric issues, like childbirth, will naturally be seen as inauthentic. One suggestion for combating accusations of inauthenticity given by these authors is that creative AI systems record details of the interactions they have in their environment, so that these observations can be referenced in future creations in an authentic way. We pick up on that suggestion here, but take it further, to encompass any system recording aspects of its existence and using them later to convey aspects of how it operates. This will help the system appear more authentic and possibly more creative, and also improve public understanding of AI and technology in general.

The Human Condition

Many disciplines and movements offer reflections on the human condition as fundamental issues of human existence and the meaning of life. Much of philosophy is dedicated to finding meaning in life, and answering questions such as how we should live, what is human nature, and what society should look like. Answers are offered in moral philosophy, utopian visions of human society, the role of truth in human enquiry and even movements such as existentialism, which suggest that ultimately life is absurd with no meaning. There is much overlap between philosophical and literary perspectives on these themes. Psychological perspectives focus more on what we need as humans, what it is to have an identity, and our personal search for meaning. Religious teachings focus on aspects such as sin, morality, cycles of life, submission before God and how to live a prudent and mindful life. Other perspectives come from biology, anthropology, history, art movements and numerous other sources.

Of particular interest here are the thoughts of Hannah Arendt (1998), who picks up on the ancient characterisation of human life into active (*vita activa*) and contemplative (*vita contemplativa*) elements, and asserts that the relative concerns are different, but neither is the more important. If we focus on the active life, Arendt builds on the ideas of Kant, Marx and others, and distinguishes human labour, work and action, and charts their changes through Western history, affected by important world events.

To make progress, we take a relatively simplistic and constrained view of the human condition, specifically grounded in the vita activa of events that actually happen to people, rather than emotions they feel or thoughts they have. Given our aim of describing a computational parallel, this seems appropriate, because events and actions do take place which involve machines in similar ways to those involving people, but the idea of machines feeling or thinking is controversial and adds extra complexity. Taking a hierarchical view, and starting at the highest level, we see the human condition as a small number of under-specified but important concepts such as growth, death, conflict, aspiration and love. Under each umbrella term, we can identify different categories for types of events to be placed into, albeit with overlap. For instance, the high-level concept of death covers event types which can be categorised into: grief, mourning, pain, loss, dying, etc. Individual events related to, say, dying, involve acts such as a murder or someone attending a funeral.

To ground the discussion, imagine someone writing a poem about a leaf falling from a tree, as an analogy with which to express their grief, given a particular event: the recent loss of a loved one. Here, the aspects of the human condition being expressed might include death and love at the highest level, and grief, loss and mourning at the second level. While readers of the poem may be aware of these notions and have personal experience thereof, the individual expression of them by the poet may be new and insightful. For instance, the poem may allude to (or guide the inference of) an analogy between the tree losing a leaf and the world losing a person. This analogy may offer something new to the reader, enabling them to understand the particular grief experienced by the poet, perhaps here highlighting personal loss as part of a global loss. We could speculate that a particular reader of the poem might believe it to be 'shallow' if the poem only communicated aspects of a leaf falling or a person dying, 'deep' if it communicated some aspects of the grief of the author, or the notion of grief in general, deeper still if it led the reader to think about the notion of death, and very deep, if the poem ultimately led the reader to consider what it means to be human.

The Machine Condition: Prerequisites

Recall that we are interested in the notion that a computational system could use creative production (of poems, visual art, games, stories, musical compositions, etc.) to communicate its experiences in a way that potentially encourages audience members to think about machine existence. We upgrade the term 'experience' to 'life experience' to convey an event of particular importance to the trajectory of existence of a particular person or machine. We take the position that an entity like a machine does not need to satisfy notions of being alive or conscious to have life experiences worthy of communication through creative expression.

In sketching out aspects of the machine condition, we ask the question: "What is it like to be a machine in the year 2020?". This mirrors the famous thought experiment asking "What is it like to be a bat?" in (Nagel 1974), which raised issues of consciousness in humanity. We note that bats evolved through natural selection, and their experiences are not as easily subjected to interrogation and experimentation as they are for machines, which were engineered, not evolved. Hence understanding machine existence may be less hindered by the difficulties proposed by Nagel, who suggested that, as we don't have things like a bat body or a bat brain (nor sonar for "seeing") we can never know what it is like to be a bat from subjective experience.

We adopt the structure imposed on the human condition above as a starting point for addressing the machine condition. That is, we determine some high-level concepts, categories of events and individual event types that could describe the most important aspects of the vita activa of machines. Before providing some initial suggestions to populate this structure, we first discuss some prerequisites about individuality and computational life experiences, air some caveats and describe some requirements and desirable properties for aspects of the machine condition. To limit our exploration, we propose a specific production pipeline whereby a machine has a life experience which fits into a category of events that could be used to portray some high-level aspect of machine existence. Details of the experience are captured at the time and used later in the creative production of an abstracted artwork which people could in principle interpret via contemplation of the machine condition. The artwork could be supplemented by a text which frames the creative act and provides additional understanding of the aspects of machine existence expressed in the artwork.

As an example, imagine a generative art program producing abstracted artworks on a laptop in a public space. People sit and watch pictures emerge on the laptop screen while the software records its internal state, network environment and aspects of its external environment (through a camera and microphone). When someone accidentally spills coffee on the laptop, this leads to the laptop suddenly shutting down. When the software is next run, its internal sensors highlight that its physical environment has changed from a laptop to a desktop computer, and its code has been copied from a repository, with slight changes imposed. This life experience could be used later on by the system to portray one of a number of aspects of machine existence, expressed via an artwork and an accompanying piece of text.

Individual Life Experiences

It is important to address the notion of a machine *self*, i.e., what exactly we mean when we say a machine has expressed *itself* creatively? We restrict ourselves initially to consideration of software and hardware intertwined into what would normally be considered one system. Such a system might simply be software on a laptop computer changing pixels on a screen, or might be a desktop computer running software that controls a robotic arm. We further specify that the system should be able to record details of events that happen to it, in order to use them in creative production later on. Both the software and hardware component can change over time, but it is reasonable to think of it as still the same system, similar to how a particular person changes physically and environmentally over time, but never has a different self.

With the term *life experience*, we leave open the full range of experiences a system could have. At this stage, we specifically include (i) changes in the code and architecture of its software (ii) changes in its hardware (iii) changes in the data it processes (iv) changes in its computational and network environment (v) changes in its physical environment (vi) interactions with other software systems and (vii) interactions with people. We assume the software can record such experiences with similar or higher accuracy than a person.

Caveats, Requirements and Properties

It may be tempting to apply a mapping from aspects of the human condition onto aspects of machine experience. For instance, we could look at the notion of death in humanity and search for machine experiences such as the deletion of it's code, onto which to map the notion of death. There are (at least) three difficulties with this approach. Firstly, aspects of machine existence which map nicely to human existence may not be particularly representative, e.g., entire deletion of code is not a particularly common thing to happen to machines, and certainly not for the type of systems we are focused on. Secondly, the analogy may not hold perfectly, and actually serve to confuse our understanding of machine existence rather than clarify it. For instance, the wholesale deletion of code is more akin to the science fiction idea of wiping a person's brain, than the complex notion of human death. Thirdly, mapping the human condition onto software existence likely serves more the purpose of understanding humanity than increasing our understanding of machines.

We have to be equally wary of mapping actual machine life experiences onto aspects of the human condition, for much the same reasons as above. A number of projects have demonstrated that software can be specifically engineered so that we can project an element of the human condition onto it. For instance, imagine a generative art system described by its author as 'suffering' when it senses that no-one is watching it. Such an exercise demonstrates only that it's possible to engineer scenarios in which actual computational situations map onto aspects of the human condition. However, the situation is in part artificial, because the relevant life experiences (i.e., the event of sensing that no-one is watching) have been engineered entirely for the project. Such projects may encourage us to think about human suffering, but as there is no machine equivalent, they do little to increase our understanding of machines.

To make some progress, we can specify some initial requirements of the life experiences which could be used in creative expression by machines as follows:

- The experience actually happened to the machine as part of its processing in a non-cyclic way, i.e., for purposes other than harvesting experiences for creative expression.

- The experience is reasonably unusual and distinctive to day-to-day experiences that the machine has.

• The experience reasonably fits into one or more categories of high-level software existence that captures an important aspect of the machine condition – with some suggestions for such categories supplied below.

Over time, we expect these requirements to be relaxed, as new ways for machines to communicate through creativity emerge. In addition, we can suggest some desirable, but not necessary, qualities of machine life experiences, to help further narrow down the scope of creative expression exhibited by initial implementations. In particular, it may be sensible in the first instance to concentrate on experiences which highlight aspects of the machine condition which are most different to the human condition. This may entail that the communication of machine existence is less easily confused with a projection of humanity, and perhaps speed up understanding of machine existence overall, i.e., by tackling differences to us rather than similarities which might be easier to understand. Also, it might be sensible to concentrate on life experiences and aspects of the machine condition which are possible to express succinctly in a single artefact, so that any explanatory text (see below) is kept to a minimum.

Aspects of Machine Existence

We have identified five high-level under-defined concepts into which machine-centric events can be categorised, namely *transience*, *learning*, *humanity*, *work* and *physicality*. These are meant to be at the same level as the notions of death, growth, aspiration, conflict, love, etc., in the human condition, and each is expanded below in terms of the categorisation of events that they afford. These areas are given in no particular order, and there is much fluidity in the categorisation, i.e., some event types could be moved between areas, or should perhaps be considered in multiple categories, as is the case with event types for the human condition.

Transience

The changing nature of basically all aspects of the existence of a particular machine deserves to be considered as a particularly important notion in describing machine existence. In no particular order, transience of machine existence includes events which fit into categories such as network variation, changes in data, alterations to the machine's physical and software components, changes in the local external (human) environment in which the machine is located, and global changes to the human world. Event types which might class as life experiences in these categories include: a substantial new module in a machine's code; the changing of a robotic component; movement of the machine to a new venue; a new trend on a social media platform like Twitter, if the machine is processing data from this stream; and web pages it refers to, or an entire network, going offline.

Returning to the generative art system in the coffee-spill example, suppose it retrieves images to use as art materials by querying sites on the internet. Such processes are liable to change regularly, due to the everyday shift in the structure of the internet. In the case of major events, like a nameserver being unreachable, this could result in dramatically different times or routes for connections, which could affect how long creative processes take, what order different tasks are completed in or how far a search can be conducted. In addition to network structure changing, the content may change, especially when searching for data that is algorithmically curated. Searching for images about 'coronavirus' in early 2020 will result in very different results depending on the month the search is conducted, where in the world the originating connection is, and what other things have been previously searched for. While people may not even perceive these shifts, software may experience it very differently, particularly if the system evaluates the results of the search in order to make a decision. In general, if a machine could communicate the transient nature of the network it operates in, this may help people to understand unexpected or seemingly inconsistent outputs or decisions from it.

As software is developed, its codebase grows and contracts as code is added, refactored or removed, and it is common to use version control to manage the codebase and record how a project develops. Analogies with the human experience such as growth, development, evolution, learning and training don't quite fit the experiences of a machine witnessing in short timescales its changing self. Recording and retelling events relating to its changing codebase and subsequent changes in its affordances offers opportunities for a machine to express the transient nature of its existence as an entity. Such changes are inaccessible to an observer, but deeply relevant to their experiences of the machine, since the code informs what the machine does, is, and can potentially achieve. If a machine can communicate aspects of the evolution of its codebase, this may help people understand the impact of bugs and how they lead to errors or delays, and/or help people be more patient with respect to missing features.

Learning

How a system comes into existence is something worthy of communicating through creative expression, as understanding the origins of a machine may help people comprehend how it works and the impact it has on their life. One major concept in this context, which acts as an umbrella to many aspects of the machine condition, is learning. Given the impact that one set of techniques in this area, deep learning, is currently having on society, and that these techniques are the main focus of XAI work, expressing elements of learning seems imperative. Categories of event types here could include: the training of a new machine learning model which forms part of a machine; a new dataset being used for the training; evaluating a trained model's predictive accuracy or it's value against some other set of measures; the running of such models for predictions, categorisations or generative tasks; and analysis of the results of such running. We can also include events which relate to the human programming of a machine, or the automatic generation of code for a machine, through, for instance, genetic programming techniques. Another set of events which might fit under the umbrella of learning include those related to physical memory (hard-drive or RAM) access, and differences in human and machine memory around permanence, ease of retrieval and representation could be expressed in relation to such events.

In the coffee-spill example, imagine the generative system

was augmented with a machine vision neural model able to analyse the art images that it produced. The first running of the machine vision model to analyse a generated image would constitute a life experience of significant magnitude. This could be conveyed through creative production in such a way as to highlight some of the above general aspects of machine learning, e.g., comparing how people see textures, colours and shapes in an image, contrasting with the machine, which calculates numerical outputs from an artificial neural network to analyse an image.

Humanity

The machine existence is decidedly within a world of human existence. Machines are often made by and for individual people and human communities. They interact with and influence people, but are not treated as equals, but rather as employees for human employers. In general, software is programmed/trained to be good at things that people want or need to do, like play chess or detect spam email. Indeed, there is an assumption in some quarters of AI research that achieving human level and human-style abilities is the only goal of the field. This all-pervasive enveloping of machine existence by humanity forms an umbrella concept to add to the description of the machine condition. Some categories of events here include: human-computer interactions, whether with programmers, users or audience members; decision making, as machines make decisions that affect people, and vice-versa; communication, where the aim is to convey some information to a person, rather than to interact with them; execution, capturing events where someone commands a machine to do something; and agency, where initiative in a scenario switches between human and machine.

Only in the last 50 years have people begun to have interactions where another entity/species (i.e., computers) has more cognitive abilities than it, albeit only in very specific ways. Machines have interactions with a more intellectual species (humanity) on a daily basis, and it may be instructive for people to understand this from a machine's perspective. Moreover, how the inequality between human and machine plays out in events that happen to a machine may be a good target for creative expression, especially if machines begin to be in charge in certain scenarios, i.e., machines begin to employ people for their benefit, not the benefit of other people. Some other issues which could be addressed include: differences in aspects of mortality/immortality between humans and machines and the difficulty machines may have in describing how they operate to people.

In the coffee-spill example, an event with a person led to a life experience for the machine. This could be used to express transiency as the software moved from laptop to desktop, but could also be used to express the relationship machines have with people, in particular the low level of agency machines tend to have: no matter how much autonomy a machine is seemingly given, it can still be rendered useless by someone, whether on purpose or by accident.

Work

Machines exist primarily to carry out work which leads to changes in a physical and/or virtual environment and/or the creation of new knowledge and artefacts via information processing. If any one aspect of the machine condition most captures the essence of machine existence, this is probably it – machines tirelessly and endlessly carry out work of value to human society. Event categories under the umbrella of work could be associated with: how software operates, e.g., loops, conditionals, subroutines, etc.; data and how it is analysed and transformed; execution traces; energy consumption; the degradation of hardware; evaluation, as people and machines determine how good a piece of machine work is; the nature, collation and storage of output; hierarchies of control and responsibility in software, hardware and human groups; and benefactors – whether people or other machines – who consume the results of machine work.

In the coffee-spill example, the generative process is clearly the work that the machine undertakes, and this is stopped not by the software, but by an accident. One aspect of difference between human and machine existence that could be expressed here is the notion of responsibility – while it is the machine undertaking the work, it holds no responsibility for whether the work gets done. In general, the output of machine work tends to be tailored for human consumption, but the processing and reasoning behind the undertaking of a piece of work does not. If machines could express other aspects of the work they do, this could reduce frustration when software performs poorly, and lead to improved computational thinking in the general public.

Physicality

Computational systems, even though they can be thought of abstractly, and aspects such as code tend not to be thought of in physical terms, do have a physical reality. This ranges from a server quietly processing information in a dark basement to a robot constructing a car in a factory. The nature of this physicality and how machines both sense and affect the physical world, is an inescapable part of the machine condition. Event categories here could be associated with: measurements, such as weight, mass, volume, degrees of freedom, location, reach, speed and accuracy; presence, where machines occupying physical space affects how people react to them; upgrades, where better parts are substituted; sensors and the nature of the data they record; the presence of noise in sensors, actuators and the environment, how it affects machine operations and how it can be dealt with; the changes that physical operations have on the environment, which could be permanent, e.g., welding one car part to another, or temporary, e.g., displaying an image on a screen.

The generative system in the coffee-spill example could use its camera to capture pictures of its surroundings as art material for making art, and it could easily be attached to a printer so that physical printouts of its work could be made. This may increase the presence that the machine has, and how people react to it, which could be expressed in later artworks. In general, machines sense and affect the world differently to people, and if they can express these differences, we will probably gain a deeper understanding of the world around us. Machine physicality expressed through creative production may also help us come to terms with the fact that we are indeed sharing the world with machines.

Creative Expression of Machine Existence

The pipeline mentioned above suggests a straightforward way in which a machine could express aspects of its existence, and that of machines in general, prompted by a life experience it had. The construction of an artefact, whether a piece of music, visual art, poem, game or otherwise, will naturally draw on the many generative techniques and methodologies developed in computational creativity research and elsewhere. We offer here some general thoughts which could supplement this wealth of knowledge.

The two events in the coffee-spill running example above (the spill and the movement of the software from laptop to desktop), could be used to tailor the generative software's art production techniques in order to convey to people aspects of the machine condition related to transiency (with the software being moved and changed), humanity (how the machine inhabits the world with clumsy people who do unpredictable things) and possibly work (how its work can be stopped by an outside force completely beyond its control). The choice of what to express will be key, and it may be sensible to limit the scope to one aspect per artefact produced, to increase clarity. The tailoring of the generative process could be achieved by the software altering its generative parameters, workflow and/or source material, but could equally be achieved by the addition of new code by the software itself or by a programmer.

One possibility would be to take a literal approach and juxtapose some imagery of a coffee spill with a picture of a laptop and a desktop computer. However, this might be criticised as not being very deep, given that it does not offer much opportunity for interpretation or lead to insights into the nature of machine existence. Another approach may be to allude in the artwork to an idea from human culture, for instance the science fiction meme of one person's mind being trapped inside the body of another, or a particular work from human culture, for instance Kafka's Metamorphosis. Given that the artworks are intended for human audiences, references to ideas and artefacts from human culture may help to communicate aspects of software existence. A third approach would be to make analogies to human existence. While we advocate focusing on events which happen to machines and not drawing strict analogies at high levels (such as death being the same as code deletion), it makes sense to reference human life in particular generated artworks. This is because the looseness of the analogy may encourage audience members to interpret the message personally and possibly grasp the ideas expressed on their own terms. It would be important here, though, to remember that the aim is to communicate aspects of machine existence through reference to human life, not the other way around.

While many people in technological circles may understand how machines operate and the environments they work in, it is fair to say that the general public do not fully know what software is and does and how it controls machines. As mentioned above, the idea of software framing its work with explanatory text about its motivations, processes and outputs could help people to project notions of creativity onto the software. Given the differences between machine and human existence, makes sense to add to the potential focal points for framing some details of the individual events (e.g., coffee spill/restart) and bigger picture concepts (e.g., transience, physicality) that influenced the conception, production and assessment of a particular artefact.

Ultimately, people consuming computer-generated artefacts will have to put in some effort to think about the concepts raised in their own terms, if they want to understand more about machine existence. There will be many aspects of this that people will find difficult to grasp, and these could be targets where more explicit framing could be employed, to supplement more abstract presentations in the generated artefacts. These difficulties will likely include issues of scale, both in terms of volume of data and rapid change. Other difficulties might involve the complexity of the processing undertaken by a machine, involving thousands of lines of code, or the black-box homogenuity of systems like artificial neural networks. Further difficulties may include counter-intuitive notions and other-worldliness which make comprehension of how a machine works hard, e.g., as in quantum computing. The opposite of other-worldliness might also cause difficulties, i.e., false equalities resulting in the projecting of notions of humanity onto machines which cause errors in understanding, e.g., thinking of a robotic arm as being the same as a human arm. In light of these difficulties, it may be necessary for the generative aspects of a machine expressing its existence to employ a model of general human comprehension of machine existence, and use this to determine which aspects of the machine condition to portray in the artefact produced and the framing text.

Conclusions and Future Work

We have argued the case for why machines able to express aspects of their existence through creative production could offer a new purpose for computational creativity systems, and drive the field forward. We provided a simplified account of the human condition which suggested a structure for a parallel understanding of the machine condition. After describing some prerequisites, we discussed event types and categories thereof which broadly fit into five areas related to machine existence, namely transience, learning, humanity, work and physicality. Finally, we provided some thoughts on creative expression of the machine condition by computational creativity systems, and overall we hope that this work provides an initial framework to refer to when implementing a creative AI system able to express through its generated artefacts what it is like to be it.

The rise of generative deep learning and other techniques in the last decade has meant that automatically producing high quality artistic artefacts in volume is rapidly becoming less of an issue in computational creativity. This provides an opportunity to concentrate on the societal issues in the field, in particular the questions of why we want machines to be creative, and in which cultural contexts this would be appropriate. Suggesting that machines be creative in order to communicate about their experiences opens up new avenues for research with respect to these existential issues. We wouldn't expect or want a ticket machine in a railway station to draw a picture or pen a poem to explain to us how the network is down and that this reflects the transiency of

machine existence. However, it might be beneficial for an household robot to provide an accessible account of an incident or decision in some artistic form.

We ask children to draw pictures to express difficult events that occurred, and these can be enlightening, taken together with more direct communicative approaches like interviews and therapy sessions. The same could be true for machines, especially if the artefacts produced make us think about an aspect of the machine's operation that we previously didn't understand. As machine creativity rises, there may be a tendency for human creative activity to be prized in society, perhaps for reasons of community and authenticity. If this is the case, and we find that machine communication through expressive art is successful in the ways described above, it is not impossible to think that computational creativity research may ultimately find most benefit in communicating the machine condition, with the value of generated artefacts a welcome side-effect.

There are many questions about the machine condition left unanswered, and many ways in which this idea could be carried forwards. Firstly, we could investigate the ways (if any) in which certain groups of people care about what it's like to be a machine. Improved understanding of how machines operate and affect society are general benefits arising from detailing and understanding the machine condition. However, there may be more specific reasons that people want to think of machines as companions, carers and collaborators, where an understanding of machine existence may help. Secondly, we could be more concrete about what a machine experience is, expanding past event-based experience and discussing how a machine would know whether it was having a life experience or not, and what other types of experience it might have. Thirdly, we could draw on semiotics and communication theory as qualitative and quantitative frameworks to distinguish what a machine tries to tell us about its existence, how this is signified through different artefacts, and how effectively and efficiently it is communicated. Finally, we could address the moral issues thrown up by people imposing notions of a machine condition, rather than them coming from the machines themselves. Indeed, one long-term milestone might be machines which define their own condition and surprise us by describing aspects we hadn't previously thought of.

Acknowledgments

We thank the anonymous reviewers for their insightful feedback. CG was funded by the Academy of Finland Flagship programme "Finnish Center for Artificial Intelligence" (FCAI) and by EPSRC grant EP/L015846/1 (IGGI CDT).

References

Arendt, H. 1998. *The Human Condition*. Uni. of Chicago Press.

Arrieta, A.; Diaz-Rodriguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R. and Herrera, F. 2020. Explainable AI: Concepts, Taxonomies, Opportunities and Challenges. *Informat. Fusion* 58.

Boden, M. 1990. *The Creative Mind*. Weidenfeld/Nicolson.

Boden, M. 2010. *Creativity & Art: Three Roads to Surprise*. OUP.

Charnley, J.; Pease, A.; and Colton, S. 2012. On the Notion of Framing in Computational Creativity. In *Proc. ICCC*.

Colton, S.; Pease, A.; and Saunders, R. 2018. Issues of Authenticity in Autonomously Creative Systems. In *Proc. ICCC*.

Colton, S. 2008. Creativity vs. the Perception of Creativity in Computational Systems. In *Proc. AAAI Spring Symp. Creat. Systems*.

Cook, M.; Colton, S.; Pease, A.; Llano, T. 2019. Framing in Computational Creativity: A Survey and Taxonomy. *Proc. ICCC*.

Corneli, J.; Jordanous, A.; Shepperd, R.; Llano, M. T.; Misztal, J.; Colton, S. and Guckelsberger, C. 2015. Poetry Workshop: Making Sense of Work in Progress. In *Proc. ICCC*.

Gervás, P. 2009. Computational Approaches to Storytelling and Creativity. *AI Magazine* 30(3).

Guckelsberger, C.; Salge, C.; and Colton, S. 2017. Addressing the 'Why?' in Computational Creativity: A Non-Anthropocentric, Minimal Model of Intentional Creative Agency. In *Proc. ICCC*.

Guzdial, M.; Reno, J.; Chen, J.; Smith, G.; and Riedl, M. 2018. Explainable PCGML via game design patterns. Proc. AIIDE Wshops.

Hantula, O. and Linkola, S. 2018. Towards Goal-aware Collaboration in Artistic Agent Societies. In *Proc. ICCC*.

Jennings, K. E. 2010. Developing Creativity: Artificial Barriers in Artificial Intelligence. *Minds & Machines* 20(4).

Jordanous, A. 2016. Four PPPPerspectives on Computational Creativity in Theory and in Practice. *Connection Science* 28(2).

Kantosalo, A. and Toivonen, H. 2016. Modes for Creative Human-Computer Collaboration. In *Proc. ICCC*.

Jordanous, A. 2019. Evaluating Evaluation: Assessing Progress and Practices in Computational Creativity Research. In *Computational Creativity*. Springer.

Linkola, S.; Kantosalo, A.; Männistö, and T.; Toivonen, H. 2017. Aspects of self-awareness: An anatomy of metacreative systems. In *Proc. ICCC*.

Llano, M. T.; d'Inverno, M.; Yee-King, M.; McCormack, J.; Ilsar, A.; Pease, A.; and Colton, S. 2020. Explainable Computational Creativity. In *Proc. ICCC*.

McCormack, J.; Gifford, T.; and Hutchings, P. 2019. Autonomy, Authenticity, Authorship and Intention in Computer Generated Art. In *Proc. 8th EvoMusArt Conf*.

Nagel, T. 1974. What Is It Like to Be a Bat? *Phil. Review* 83(4).

Nguyen, A.; Clune, J.; Bengio, Y.; Dosovitskiy, A.; and Yosinski, J. 2017. Plug and play generative networks: Conditional iterative generation of images in latent space. In *Proc. CVPR*.

Pérez y Pérez, Rafael. 2018. The Computational Creativity Continuum. In *Proc. ICCC*.

Ritchie, G. 2019. The Evaluation of Creative Systems. In *Computational Creativity*. Springer. 159–194.

Saunders, R., and Gero, J. S. 2001. Artificial Creativity: A Synthetic Approach to the Study of Creative Behaviour. In *Computational and Cognitive Models of Creative Design V*.

Saunders, R. 2011. Artificial Creative Systems and the Evolution of Language. In *Proc. ICCC*.

Saunders, R. 2012. Towards Autonomous Creative Systems: A Computational Approach. *Cognitive Computation* 4(3).

Veale, T.; Cardoso, F. A.; and Pérez y Pérez, R. 2019. Systematizing Creativity: A Computational View. In *Computational Creativity*. Springer.

Ventura, D. 2016. Mere Generation: Essential Barometer or Dated Concept? In *Proc. ICCC*.

Ventura, D. 2017. How to Build a CC System. In *Proc. ICCC*.

Wiggins, G. A. 2006. Searching for Computational Creativity. *New Generation Computing* 24(3).

Zeiler, M. D., and Fergus, R. 2014. Visualizing and understanding convolutional networks. In *Proc. ECCV*.

Zhu, J.; Liapis, A.; Risi, S.; Bidarra, R.; and Youngblood, G. M. 2018. Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In *Proc. IEEE CIG*.