Proceedings of the 7th Computational Creativity Symposium at AISB 2021 (CC2021)

At the AISB 2021 Convention
Online — 9 April 2021

Juan Alvarado and Anna Jordanous (Editors)

https://sites.google.com/view/aisb2020cc/

Program Committee

Organising Committee

Juan Manuel Alvarado López [j.alvaradolopez@qmul.ac.uk]
Anna Jordanous [A.K.Jordanous@kent.ac.uk]

PC Members

Maya Ackerman(Santa Clara University)
Amílcar Cardoso (University of Coimbra)
Simon Colton (Queen Mary University of London)
Mark d'Inverno (Goldsmiths, University of London)
Jeremy Gow (Queen Mary University of London)
Colin Johnson (University of Kent)
Anna Kantosalo (Aalto University)
Carlos León(Universidad Complutense de Madrid)
Jon McCormack (Monash University)
Hannu Toivonen (University of Helsinki)
Tony Veale (University College Dublin)
Dan Ventura (Brigham Young University)
Geraint Wiggins (Vrije Universiteit Brussel / Queen Mary University of London)

Preface

This volume contains the papers presented at 7th Computational Creativity Symposium at AISB 2021 (CC2021) held online on April 9th, 2021. The conference was hosted at the AISB 2021 Convention for the Society for the Study of Artificial Intelligence and Simulation of Behaviour Originally this symposium and the Convention were scheduled for 2020, but were postponed to 2021 following COVID-19 and associated lockdowns.

The symposium features a number of presentations covering a range of topics in the evolving field of Computational Creativity. Issues addressed included practical work in the area, theoretical approaches to creativity, and philosophical questions raised on the potential of non-human creative agents.

Over the last few decades, computational creativity has attracted an increasing number of researchers from both arts and science backgrounds, from academia and industry. Philosophers, cognitive psychologists, computer scientists and artists have all contributed to and enriched this area of research.

Many argue a machine is creative if it simulates or replicates human creativity (e.g. evaluation of AI systems via a Turing-style test), while others have conceived of computational creativity as an inherently different discipline, where computer generated (art)work should not be judged on the same terms, i.e. as being necessarily producible by a human artist, or having similar attributes, etc.

This symposium aims at bringing together researchers to discuss recent technical and philosophical developments in the field, and the impact of this research on the future of our relationship with computers and the way we perceive them: at the individual level where we interact with the machines, the social level where we interact with each other via computers, or even with machines interacting with each other.

This year we were delighted to also run a Show-and-Tell demo session as well as the paper presentations, showcasing demonstrations of computational creativity research results as well as more traditional talk presentations.

Juan Alvarado and Anna Jordanous (Organising Committee)
March 2021

Table of Contents

Papers

Show and Tell Extended Abstracts

Assessing Creativity of MEXICA: An Application of Ritchie's Criteria

Juan Alvarado[1] and **Geraint A. Wiggins**[2]

Abstract. We use Ritchie's criteria for the evaluation of creative systems to analyse MEXICA. Ritchie's criteria are for humans to use, but in this analysis, we are using them so the system can test itself. We do this analysis to get information about MEXICA's performance. With this analysis, we can delve into how MEXICA is exploring its conceptual space. MEXICA could improve its performance by changing the execution parameters it uses. We can repeat this analysis and we would optimise MEXICA's result.

1 INTRODUCTION

When analysing computer systems (non-creative as well) we have to consider several factors. For example, the inner process of the system, inputs and outputs. If we have expectations on how the system should work or what the outputs should be, the analysis must consider this too.

Boden [1] suggests conceptual spaces. She points out that creative ideas exist there. She suggests they have an origin in the culture of the creators, and they are any disciplined way of thinking familiar to (and valued by) a social group.

Boden [1] argues we can find concepts in a CS by *Combination*, *Exploration* and *Transformation*. She states that by combination it would be possible to pick out two ideas and put them alongside each other. We should combine ideas based on existing unnoticed links between them. It should also exist a database of ideas to put them together. She states that by exploring a CS someone may see concepts not discovered yet. There must be rules to explore the space, allowing us to find new concepts by using these known rules. She also points out transformation as a third strategy to find new concepts. By transformation, the form of CS changes because the rules have changed. We may find different concepts because the space available for us to find them is different.

Wiggins [11] presents a proposal to formalise Boden's ideas on creativity. He argues that at first sight Boden proposal lacks elements to use it consistently. He formalised the concepts in Boden's theory so they can be better applied and understood.

Following Wiggins [11], there are sets of rules $\mathcal{R}$ to define the conceptual space, $\mathcal{T}$ to traverse it, and $\mathcal{E}$ to evaluate concepts. There is also an interpretation function which has access to all those sets of rules, and given an initial set of concepts, it can produce a set of output concepts.

This is good because we can test Boden's [1] ideas and explore the conceptual space following the rules of the creative agent. We

can test how the agent is traversing the space ($\mathcal{T}$), and we can test whether it is exploring inside the CS ($\mathcal{R}$) or outside it. We can test concepts ($\mathcal{E}$) even those outside the conceptual space. We can also try transformational creativity by changing the rules of the different sets.

Creative systems are expected to produce original and valued results. They should also relate, to some extent, to the domain for which they have been generated, i.e. when the outputs are jokes and it was the purpose of the program to produce jokes. Boden [1] also argues that creative ideas must be original. Jennings [3] points out that one important characteristic of an autonomous creative system is the non-randomness. Ventura [10] also highlights a creative system must have intentionality. So, we do not expect creative systems to generate random outputs, because they could be original, but not valued. To produce valued results systems can use some appropriate criteria for the current problem or domain.

Ritchie [7] proposes some criteria for the evaluation of creative systems based on the outputs they produce. He suggests a neutral characterisation so that any result in any field can be evaluated with the same set of criteria. For Ritchie [7] the process by which a system produces an output is not a relevant part of the evaluation, because we can not observe that process and we do not make judgements considering it.

Ritchie's original criteria were intended to be used by human evaluators. For this work, however, we use a computer system to analyse MEXICA's output using the self-evaluation result of MEXICA. For the future work we could include human evaluators using the same criteria to see if they agree with this assessment. It is important to note that even when Ritchie's criteria does not take into account the inner process of the system; we are using it here to get the reasons for the results.

Using Ritchie's criteria, we can evaluate MEXICA's outputs. This way, we can get an understanding of the characteristics of the output MEXICA produces. Other works in the past have used these criteria to test the performance of systems [4, 2]

By using the ideas proposed by Wiggins [11], we can get the rules MEXICA applies to generate its conceptual space ($\mathcal{R}$), search strategy ($\mathcal{T}$) and evaluation ($\mathcal{E}$) of concepts. Combining those rules and the result of applying Ritchie's criteria [7] to MEXICA's outputs, we can have a more complete idea on MEXICA's performance. This can help us analyse the way MEXICA explores its conceptual space.

We can use Ritchie's criteria, which is meant to be used by humans, to evaluate a system by a system. Ritchie [7] points out that "these criteria take a third (external) viewpoint, in which one treats the program as an input/output data conveyor and attempts to say more precisely how it has performed."

Ritchie also points out that typicality and quality mappings are

---
[1] Queen Mary University of London, UK, email: j.alvaradolopez@qmul.ac.uk
[2] Vrije Universiteit Brussel, Belgium & Queen Mary University of London, UK, email: geraint.wiggins@vub.be

not proposed as components of the program or system (in this case MEXICA). But instead, they are measures that can be applied, for example, by having an external program judge the outcome of the program.

Using this, we could change the execution parameters of MEXICA and check, using the same combined evaluation, whether the outputs are better in terms of Ritchie's criteria. In this way, we could optimise MEXICA's performance.

The goal here is meta level evaluation. Our aim is to take these components with an external system. This would mean that we are evaluating the object level, and choosing a new version of MEXICA, because the evaluation of MEXICA the way it was working was not appropriate.

For the moment, in this paper, we are only reporting the results of applying Ritchie's criteria to test MEXICA in combination with the inner process. For the future work, we will include this evaluation to change MEXICA's parameters of execution and hence, its performance.

2 RITCHIE'S CRITERIA FOR THE EVALUATION OF CREATIVITY

Ritchie [7] developed a set of criteria, that (as he explains), if observed in the output of a system, that output should be deemed creative. A central assumption of his proposal is that any formal definition of creativity must be based on its ordinary usage, it must be *natural* and it must be based on *human behaviour* [7]. By "natural", he means that any technical definition of "creativity" which is to be used in discussing the behaviour of computer programs must capture fairly accurately the original ordinary language use of the term. He mentions "human behaviour" because of the way the word "creative" is ordinarily used when talking about (human) creativity, this should be considered with non-machine creativity as well [7].

Ritchie [7] also points out that all requirements of creativity must be observed empirically, in the same way, and with the same judgements we make for evaluating tasks performed by humans. Ritchie remarks that the process by which a system produces an output is not a relevant part of the evaluation. He argues that this is because the formal definition of achieving creativity in computer systems mimics judgements of humans, and then it should be based only on comparably observable factors, without adding extra information about the internal workings of the computer program. Ritchie [7] suggest that arguing that the inner workings of a computer program are critical in deciding its creativity move us away from the way human creativity is normally judged.

Ritchie [7] explains that creativity depends on some essential properties that must be identified and quantified as they impact the assessment, and should be present in creative products. These properties are:

Novelty To what extent is the produced item dissimilar to existing examples of its genre?
Quality To what extent is the produced item a high quality example of its genre?
Typicality To what extent is the produced item an example of the artefact class in question?

Ritchie [7] also considers that the program is influenced by a subset of basic items that he calls the inspiring set. This set could be all the artefacts known to the program designer, or items which the program is designed to replicate, or a knowledge base of known examples which drives the computation within the program. Ritchie

[7] point out he includes the inspiring set because creativity could be viewed as depending on the extent to which the program replicates the instances which guided its design.

Ritchie [7] combines Typicality, Quality, the Inspiring set I and the set of outputs of the program R in several ways to get criteria to assess whether a system has been creative. By using the notation in Table 1, where he uses the Typicality and Quality functions $(typ(x), val(x))$, Ritchie identifies sets of items falling in a range of typicality and quality and the average and relative sizes of these sets to explain the creativity of the system based on these data. The complete criteria can be found in [7].

Table 1. Ritchie's notation to explain the criteria.

The subset of X falling in a range of typicality.
$$T_{\alpha,\beta}(X) \stackrel{def}{=} \{x \in X \mid \alpha \leq typ(x) \leq \beta\}$$

The subset of X falling in a range of quality.
$$V_{\alpha,\beta}(X) \stackrel{def}{=} \{x \in X \mid \alpha \leq val(x) \leq \beta\}$$

The average value of a function F across finite set X.
$$AV(F,X) \stackrel{def}{=} \frac{\sum_{x \in X} F(X)}{|X|}$$

The relative sizes of two finite sets $X, Y, |Y| \neq 0$.
$$ratio(X,Y) \stackrel{def}{=} \frac{|X|}{|Y|}$$

3 MEXICA

Sharples [9] proposes a creative account of writing as a creative design where he explores creative writing as a process where the writer generates new material by imposing appropriate (internal and external) constraints, which can be a combination of external resources and the writer's knowledge and experience.

According to Sharples, by imposing constraints on a generative system, it is possible to form what Boden [1] describes as a conceptual space. Sharples [9] argues that creativity in writing occurs through a mutually supportive cycle of *Engagement* and *Reflection*, both guided by constraints.

Based on ideas exposed by Sharples, Pérez y Pérez [5] presents the computer model E-R and implements it in MEXICA, which has many modifiable parameters to experiment with creating a new story plot and one of its goals is to produce novel and appropriate short stories as a result of an Engagement-Reflection cycle, guided by sets of internal and external constraints, and without the use of pre-defined story-structures.

MEXICA needs two inputs provided by the user: a set of Primitive Actions (PA) and a set of Previous Stories (PS). The first gives the system the knowledge of everything that is possible to happen in a story, and the second are examples of stories (built with PA), that the system will use to build new ones.

In MEXICA a story is a sequence of events or actions which are coherent and interesting. An action is an event in a story in which characters can take part. An action has pre-conditions and post-conditions, useful to give coherence to a story and to know the consequences of the execution of an action, respectively.

During Engagement, MEXICA does not verify if the story actions satisfy pre-conditions. At this stage, as explained in [5], Engagement might produce a sequence of actions with unsatisfied pre-conditions (potentially non-coherent stories). But it might be the case that the

sequence of actions is actually coherent. So, Engagement can produce coherent and non-coherent stories.

In contrast with Engagement, Reflection verifies pre-conditions for each action in the story in progress to produce a coherent story.

The conceptual space of Reflection is a subset of that of Engagement as the first can contain only coherent stories and the second can contain coherent and non-coherent stories. This an interesting property in MEXICA because that means that Engagement and Reflection explore different conceptual spaces. We can say that the conceptual space of Engagement is bigger that the one for Reflection. This is good because there may be more concepts to find. One problem with this can be the evaluation process of MEXICA, because it takes into account the pre-conditions fulfilled, but concepts in the conceptual space of Engagement might not have them all fulfilled and that will reduce the overall score of quality for that concept.

Also, in Reflection, MEXICA considers that an interesting story includes degradation-improvement processes. When it discovers that the story in progress does not increment tension, guidelines are established in a way that the next execution of Engagement will favour retrieving actions able to produce tension to continue the story.

Boden [1] suggests that novelty is one important characteristic of creative acts. Novelty is also considered in MEXICA, and during Reflection, there are rules to assess novelty.

MEXICA verifies if the material produced during the Engaged state resembles too much any of the tales in the set of PS. The system has a parameter, called Novelty-Percentage, that determines the maximum percentage of similarity allowed between two tales; if it is exceeded, the guidelines are established to get a more original sequence of events during Engagement. This percentage has a default value of 50% and is modifiable by the user [5].

MEXICA generates, as a result of its execution, a story and several output files. The output files include information about the parameters used during the execution, the selection process for the actions to continue the story in progress, and results of the evaluation of the story, such as originality, overall evaluation, percentage of preconditions fulfilled, appropriate opening and closing, etc.

4 APPLYING THE CRITERIA TO MEXICA'S OUTPUT

Ritchie's criteria [7] use typicality, quality and novelty values to assess creativity. For MEXICA there are some particular characteristics which define what stories are (typicality) and what make them high quality and novel. From the outputs produced by MEXICA we can extract information about the story generated. We can use this information to evaluate its performance using Ritchie's criteria.

Typicality represents to what extent the produced item is an example of the artefact class. MEXICA does not report a value of typicality. It produces output files with several results related to the class of artefacts it generates. The class of artefacts MEXICA produces are short stories. Following Pérez y Pérez [5], for MEXICA, *a story is a coherent and interesting sequence of actions*. To be coherent, all preconditions must be satisfied. To be interesting, it must include a degradation-improvement process (conflict, complication and resolution) [5].

In the output files, MEXICA reports the evaluation of preconditions satisfied. It also reports the opening, closure and amplitude of the main peak, they have to do with the degradation-improvement process. Preconditions satisfied, opening, closing and amplitude of the main peak are properties of the story that get a numerical assessment value in the range 0-1. To get the typicality value for the

generated story in MEXICA, we get the average value of those properties.

This decision was made taking into account the properties pointed out in [5] a story must-have. For example, we could remove the coherence and interesting-ness properties, so a story could be defined simply as a sequence of actions. The problem with this approach is that all stories produced by MEXICA would be typical examples. Untypical examples, which are also interesting to analyse, would not exist.

Quality has to do with the extent the produced item is a high-quality example of its genre [7]. MEXICA [5] evaluates the generated story and gives an overall quality value. For this value, it considers the same properties that make a story a story, but it also considers other elements such as novelty, the number of original sequences and relationship between actions in the story.

Novelty measures to what extent the produced item is dissimilar to existing examples of its genre [7]. In MEXICA [5] the user provides the system with two files; the file of primitive actions (those that can happen in a story) and the file of previous stories (built with primitive actions). MEXICA uses the previous stories to generate new ones. It finds patterns in previous stories to continue the story in progress. MEXICA tries not to copy previous stories, but this can happen to some extent. MEXICA reports an originality value which shows how similar the generated story is to the set of previous stories (not only one story).

To apply Ritchie's criteria, the set of twenty stories in [6] have been used as MEXICA's input (the set of previous stories). Eighty-nine stories were generated and from them, typicality, quality and novelty have been extracted to apply the criteria.

Ritchie's criteria [7] include several expressions which use the *Inspiring* set. This is how Ritchie calls the examples known by the system. Here the inspiring set is the set of previous stories.

For criteria 9-10a in Table 5 Ritchie uses the expression:

$$I \cap R$$

Where I is the inspiring set and R is the set of outputs. This is the intersection of these two sets and includes elements in the output that have been replicated from the inspiring set.

When MEXICA generates a story, it can detect a previous story is being copied. If this happens, it tries to avoid using actions from the copied story to continue the story in progress.

- If no action can be added to the story in progress it gives up. At this point, $I \cap R$ is empty, because there are not elements in common in I and R.
- If an action can be added to the story in progress, then MEXICA is not copying a previous story anymore and the process continues. $I \cap R$ is empty, because there are not elements in common in I and R.

So, MEXICA does not copy previous stories but it analyses generated stories and gives an originality value to those stories.

MEXICA checks the story in progress (and the final story) against all previous stories. It can find a previous story has been partially copied but does not report which one it was. It checks if similarities can be found with more than one story. This is normal and can happen because MEXICA uses previous stories to generate new ones. Parts of different previous stories might be present in the current one. MEXICA reports an originality value based on similarities found with different previous stories (not only one but it does not report which ones). So, the intersection $I \cap R$ becomes something differ-

ent, more like a fuzzy intersection and I and R would be fuzzy sets too. This is not explored here and will be part of the future work.

For this experiment, we are considering the originality value reported from MEXICA to calculate the intersection $I \cap R$.

Ritchie [7] also uses an expression where he includes elements in the output R but not in the inspiring set I. Criteria 11 to 18, shown in Table 6, use the expression

$$R - I$$

In MEXICA, each output in R is evaluated, and it gets a value of originality. From the information extracted from the outputs files, we can say which ones have low values of originality. Because of this, we know they are similar to one or more stories of the set of previous stories. Finally we can get the difference $R - I$

Again, this is not a normal set operation. We are not checking whether the generated stories are also in the inspiring set or not. We are checking only if the originality value of the generated stories is high enough to be considered novel. If it is high enough, then they are part of $R - I$

We can proceed for novelty, the same way Ritchie does with typicality and quality. The new expression for novelty is shown in Table 2. This one extends the notation Ritchie uses. We can get a subset of R falling in a range of novelty.

Table 2. Extension to Ritchie's notation to include Novelty.

The subset of X falling in a range of novelty.
$$N_{\alpha,\beta}(X) \stackrel{def}{=} \{x \in X \mid \alpha \leq nov(x) \leq \beta\}$$

By using the expression for novelty shown in Table 2, we can get the subset of R falling in a range of novelty. This way, we can identify the elements in R from a suitable α to 1. We can call this *the original output*. This is $R - I$

We know the elements in R and the elements in $R - I$. We can define the intersection $I \cap R$ in terms of the other two. It would be the set $R - (R - I)$

This is the intersection between the output R and the input I. This intersection might not be made of exact copies of previous stories. Elements in this intersection can be stories with long sequences copied from different previous stories. For Ritchie's criteria, one important aspect of this intersection is its size.

We have calculated the intersection and difference above considering only the elements in R. We can say this intersection means something like: 'these stories in the output are too similar to one or several stories in the set of previous stories that they can not be considered novel'.

Ritchie [7] explains that for the notation of typicality and quality, we should find suitable α and β values to apply typicality and quality functions. Now, for this new novelty function proposed, we should do the same. Ritchie points out that for typicality and quality this is not a trivial task, and it is difficult for novelty too.

Table 3 shows the result of the application of Ritchie's criteria to MEXICA's output.

5 ANALYSIS

Ritchie [7] points out that there is a threshold θ for the criteria. This threshold is not easy or trivial to define. The values shown in Table

Table 3. Results of the criteria applied to MEXICA alpha=0.5, beta=0.5, gamma=0.5.

Criteria	$\alpha=\beta=\gamma=0.5$	$\alpha=\gamma=0.7, \beta=0.3$	$\alpha=\gamma=0.9, \beta=0.1$
criterion 1	0.727154472	0.727154472	0.727154472
criterion 2	0.887640449	0.662921348	0.348314607
criterion 3	0.304626311	0.304626311	0.304626311
criterion 4	0.292134831	0.168539326	0.101123596
criterion 5	0.329113924	0.254237288	0.258064516
criterion 6	0	0	0
criterion 7	0	0	0
criterion 8	0	0	0
criterion 8a	0	0	0
criterion 9	0.8	0.8	0.8
criterion 10	5.5625	5.5625	5.5625
criterion 10a	0.820224719	0.820224719	0.820224719
criterion 11	0.796790497	0.796790497	0.796790497
criterion 12	0.371393721	0.371393721	0.371393721
criterion 13	0.764044944	0.651685393	0.348314607
criterion 14	0.292134831	0.168539326	0.101123596
criterion 15	0.931506849	0.794520548	0.424657534
criterion 16	0.356164384	0.205479452	0.123287671
criterion 17	0.356164384	0.205479452	0.109589041
criterion 18	0	0	0

3 have to be tested against an appropriate θ value. This way we get a final result of the criteria explaining if the evaluation of the output yields a positive result. Here, we are not defining θ but we will check the values we have got for MEXICA using these criteria and we will make some conclusions.

Table 4 shows criteria 1-8a. Criterion 1 shows the average typicality that has been found in the output. For this criterion, we can see that there is no difference in the average typicality when parameters change because this measurement does not use parameters. Typicality values are extracted from each element in the output set, and the average is calculated. The value for this criterion suggests that, on average, the output produced by MEXICA falls into the typical category if we take a θ threshold of 0.7.

Criterion 2 gives us the proportion of typical examples in the output set. For this criterion we use α which shows the lower limit of typicality, the upper limit is 1. When α is 0.5 most of the output is considered typical. The value for this criterion drops if α gets bigger. When α is 0.9, we consider the range 0.9-1.0 (very typical examples), the criterion gives us a result of 0.348314607. This means that about a third of the output is very typical. When α is 0.7, about two-thirds of the output are typical. This makes sense with the first criterion and shows that MEXICA generate many typical examples.

Criterion 3 shows the average quality that has been found in the output. This criterion does not use parameters. Quality values are extracted from each element in the output set, and the average is calculated. This value of quality shows that many of the stories generated are of low quality.

Criterion 4 calculates the proportion of high-quality artefacts in the output. This criterion uses γ as the lower limit and the upper limit is 1. We can see that the values for this criterion decrease as γ increases. The proportion, in any case, is not large and when the γ parameter is more strict, the proportion of very high-quality artefact is very low.

Criterion 5 shows the proportion of artefacts that are typical and high-quality against the subset of typical examples in the output. Criterion 2 results in Table 3 shows the proportion of typical examples in the output. Having 89 stories in the output for this experiment and the parameters α and γ are 0.9, we get 31 typical stories. From those typical stories, the ones that are high quality are 8. The proportion is

4

low.

For MEXICA, criteria 6-8a get a result 0. This happens because these criteria use the following expression, which they combine in different ways.

$$V_{\gamma,1}(R) \cap T_{0,\beta}(R)$$

With this, we are looking for the intersection of high-quality and non-typical artefacts. This is an interesting set of criteria because it explores the possibility of finding high-quality items among the items that don't belong to the class of artefacts the system produces. We can also say the system is exploring out of the boundaries of its conceptual space, and there might be good examples. This concept is also pointed out by Wiggins [11] and he calls it an aberration.

In MEXICA's output, there is a subset of non-typical items and a subset of high-quality items, but there are no elements in common in these subsets. This happens because the evaluation function MEXICA uses takes into account properties of typical examples, and therefore a high-quality item can not be non-typical at the same time.

Criteria 6-8a look for the proportion of the intersection of high-quality and non-typical items against 1) all outputs, 2) non-typical outputs, 3) high-quality and typical, and 4) high-quality outputs. All these proportions are zero.

Table 4. Ritchie's criteria 1 to 8a

#	Criterion/Explanation
1	$AV(typ, R) > \theta$ The average typicality of artefacts created
2	$ratio(T_{\alpha,1}(R), R) > \theta$ From all artefacts created, how many are typical?
3	$AV(val, R) > \theta$ The average value of artefacts created
4	$ratio(V_{\gamma,1}(R), R) > \theta$ From all artefacts created, how many are good?
5	$ratio(V_{\gamma,1}(R) \cap T_{\alpha,1}(R), T_{\alpha,1}(R)) > \theta$ From the typical artefacts created, how many are good?
6	$ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), R) > \theta$ From all artefacts created, how many are good and non-typical?
7	$ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), T_{0,\beta}(R)) > \theta$ From the non-typical artefacts created, how many are good?
8	$ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), V_{\gamma,1}(R) \cap T_{\alpha,1}(R)) > \theta$ Comparison between good, non-typical and good, typical artefacts
8a	$ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), V_{\gamma,1}(R)) > \theta$ From the good artefacts created, how many are non-typical?

Criteria 9-10a use the intersection $I \cap R$ and the result for these criteria do not change when the parameters change because α, β and γ are not used. These criteria only use the sizes of I, R and $I \cap R$

As we discussed in the previous section, the intersection between I and R is difficult to calculate because MEXICA does not copy previous stories, but as it uses them to build new ones, they can contain sequences of them. MEXICA analyses output stories and gives them a value of originality that has to do with the sequences (and their length) have been copied rather than a single story being copied. The higher the original value, the more original the story. The range of this originality value goes from 0 to 1.

For this experiment, the threshold of originality 0.1 has been used. It means that if a story has a value of originality greater than 0.1 and less than or equal than 1.0, it is classified as original. If a story has a

value less than 0.1, it is classified as not original. We have chosen 0.1 as the originality threshold because as it can be inferred, MEXICA takes parts of other stories, so the original value is most times a high value because sequences are likely to be found in generated stories.

There are 20 previous stories. If we chose an originality threshold bigger than 0.1 and we use the novelty function in Table 2 with the output, we would end up with an intersection of 30, 40, or more elements. If they are the intersection, I (the previous stories) should have the same elements, but this is not possible because they are only 20 stories. With the threshold selected, the size of the intersection is 16.

Table 5 shows the criteria 9 to 10a. For criterion 9, it means that a high proportion of examples have been replicated from I. Ritchie [7] points out that this can be an achievement. If the system has produced nothing original but has 'merely' shown a computational route by which many interesting (known) concepts could in principle be reached, that can be a useful finding [7].

Criterion 10 has to do with producing more than just the inspiring set and this, as explained by Ritchie [7], can be seen as a symptom of creativity. Here, for MEXICA, it means that it produces 5+ times more than the things it copies.

Criterion 10a is a revision of criterion 10 to reflect the same thing and avoid undesirable properties of criterion 10. For example, if the intersection is empty, it yields a division by zero.

Table 5. Criteria including $I \cap R$

#	Criterion/Explanation
9	$ratio(I \cap R, I) > \theta$ Proportion of examples that have been replicated from the inspiring set
10	$ratio(R, I \cap R) > \theta$ Proportion of artefacts created different from those in the inspiring set
10a	$(1 - ratio(I \cap R, R)) > \theta$ Proportion of all those that have not been copied from I in the set of artefacts created

For criteria, 11-18 Ritchie uses $R - I$. For this experiment, this set is also built using the originality threshold 0.1 and the novelty expression in Table 2. Elements in the output R with an original value bigger than 0.1 are considered different from examples in the inspiring set I. For these criteria, we consider the elements in $R - I$, which are those artefacts in the output but different from those in I. Following the selection process described for this experiment, the size of $R - I$ is 73.

Table 6 shows criteria 11-18, all related to expression $R - I$. Criterion 11 gives us the average typicality value of elements of $R - I$. This value shows that elements in the output and different from I are very typical. It is related and similar to criterion 1.

Criterion 12 is similar to the previous, but this one is for quality. It shows that the quality of elements in $R - I$ on average is not high.

Criterion 13 measures the proportion of typical examples in $R - I$ against R. This calculation depends on the value of α, the bigger the value, the more strict the typicality. The result for this criterion when α is large is a little low; a third of the elements in $R - I$ is very-typical. However, the result of this criterion goes up very quickly when α lowers its value. This means that with moderate values of typicality, MEXICA has a good performance.

Criterion 14 is similar to criterion 13, but this one is for quality. Here, the result is not good for all variations of γ and it reflects a

similar notion than criterion 4.

Criterion 15 is similar to criterion 13. It compares typical artefacts, but in this case, the proportion of typical artefacts in $R - I$ is compared against $R - I$ itself. With this criterion, we analyse $R - I$ in isolation. Again, MEXICA do well for this criterion which involves typicality.

Criterion 16 is similar to criterion 14. It tests the proportion of quality artefacts in $R - I$ against $R - I$ itself as in the previous criterion. We analyse the quality of the outputs of $R - I$ in isolation. The values here are not high.

Criterion 17 measures the proportion of high-quality and high-typicality elements in $R - I$ against all element in $R - I$. This is kind of similar to criterion 5 but we only consider $R - I$. The results are not the same, but they differ little.

Criterion 18 uses non-typicality and high-quality in $R - I$. As it happened for criteria 6-8a, the intersection is empty and the result is zero.

Table 6. Criteria including $R - I$

#	Criterion/Explanation
11	$AV(typ, (R-I)) > \theta$ The average typicality of the artefacts created that have not been replicated.
12	$AV(val, (R-I)) > \theta$ The average value of the artefacts created that have not been replicated.
13	$ratio(T_{\alpha,1}(R-I), R) > \theta$ From all artefacts created, how many, which have not been replicated, are typical?
14	$ratio(V_{\gamma,1}(R-I), R) > \theta$ From all artefacts created, how many, which have not been replicated, are good?
15	$ratio(T_{\alpha,1}(R-I), (R-I)) > \theta$ From the artefacts created that have not been replicated, how many are typical?
16	$ratio(V_{\gamma,1}(R-I), (R-I)) > \theta$ From the artefacts created that have not been replicated, how many are good?
17	$ratio(V_{\gamma,1}(R-I) \cap T_{\alpha,1}(R-I), (R-I)) > \theta$ From the artefacts created that have not been replicated, how many are typical and good?
18	$ratio(V_{\gamma,1}(R-I) \cap T_{0,\beta}(R-I), (R-I)) > \theta$ From the artefacts created that have not been replicated, how many are non-typical and good?

6 CONCLUSIONS

After this analysis, we have seen some characteristics of MEXICA's output. We can say that MEXICA do well with typical artefacts. It has good mechanisms to verify stories and correct them if problems arise. MEXICA tries every time to produce a story under the standards it has established. This is a good thing because the outputs it produces are consistently typical stories.

It is because of MEXICA has a very good performance with typicality that we can see a problem. Criteria 6-8a got a zero result. This is because these criteria look for untypical and high-quality artefacts but MEXICA does not evaluate untypical artefacts as high-quality therefore the intersection $V_{\gamma,1}(R) \cap T_{0,\beta}(R)$ is empty. The evaluation of untypical artefacts is an interesting improvement MEXICA could have. It would give the possibility to explore beyond the boundaries of its conceptual space and still be able to evaluate the result. It could change its conceptual space to include the space of untypical but high-quality concepts.

We should also note that criteria related to the quality of outputs got bad results. This can happen because the evaluation rules of the system are too strict. Also because they consider some specific factors in stories. It could also happen because of the knowledge MEXICA has; primitive actions or previous stories. Information extracted from these inputs is also used to evaluate quality. And also this might happen because of the parameters it uses to evaluate the story.

This experiment was carried out to get an evaluation of MEXICA's output and find out what aspects can be improved when operating it. If we know that the evaluation quality or novelty is an issue, we can try to change the parameters of the system and generate new outputs. With these new outputs we can apply Ritchie's criteria again. We can compare previous results and modify parameters of the system to continue to operate MEXICA optimising the output each time.

There is room for further discussion for this experiment. We pointed out the intersection $I \cap R$ can be treated as a fuzzy intersection. For that we could also treat all sets as fuzzy sets. This has been discussed but not elaborated by Ritchie [8]. The α parameter for the novelty function (0.1) can be changed to see what are the results. This will continue to move us in the direction of fuzzy sets.

In MEXICA, the definition of a story can be changed so it can include a conceptual space different than the one for coherent and interesting stories, or these concepts could be redefined.

Also, as part of the future work, we should apply all this analysis to Wiggins [11] ideas and then optimise MEXICA's performance.

REFERENCES

[1] Margaret A. Boden, *The Creative Mind: Myths and Mechanisms*, Abacus, 1990.

[2] Pablo Gervás, 'Exploring quantitative evaluations of the creativity of automatic poets', in *Workshop on Creative Systems, Approaches to Creativity in Artificial Intelligence and Cognitive Science, 15th European Conference on Artificial Intelligence*, (2002).

[3] Kyle E Jennings, 'Developing Creativity: Artificial Barriers in Artificial Intelligence', *Minds and Machines*, **20**(4), 489–501, (2010).

[4] Francisco Câmara Pereira, Mateus Mendes, Pablo Gervás, and Amílcar Cardoso, 'Experiments with assessment of creative systems: An application of ritchie's criteria', in *Second Computational Creativity Workshop, IJCAI2005, July 2005*, Edinburgh, Scotland, (07/2005 2005). School of Informatics, University of Edinburgh, School of Informatics, University of Edinburgh.

[5] Rafael Pérez y Pérez, *MEXICA : A Computer Model of Creativity in Writing*, Ph.D. dissertation, The University of Sussex, 1999.

[6] Rafael Pérez y Pérez, *Mexica: 20 Years-20 Stories [20 años-20 historias]*, Counterpath Press, Denver, Colorado, 2017.

[7] Graeme Ritchie, 'Some empirical criteria for attributing creativity to a computer program', *Minds and Machines*, **17**(1), 67–99, (2007).

[8] Graeme Ritchie, 'A closer look at creativity as search', in *International Conference on Computational Creativity*, pp. 41–48, (2012).

[9] Mike Sharples, 'An account of writing as creative design', *The science of writing: Theories, methods, individual differences, and applications.*, (January), 127–148, (1996).

[10] Dan Ventura, 'How to Build a CC System', in *Eighth International Conference on Computational Creativity (ICCC'17)*, eds., Ashok Goel, Anna Jordanous, and Alison Pease, pp. 253–260, Atlanta, GA, (2017). ACC.

[11] Geraint A. Wiggins, 'A preliminary framework for description, analysis and comparison of creative systems', *Knowledge-Based Systems*, **19**(7), 449–458, (2006).

First Experiments in the Automatic Generation of Pseudo-Profound Pseudo-Bullshit Image Titles

Simon Colton,[1,2]   **Sebastian Berns**[2] and **Blanca Pérez Ferrer**[3]

Abstract. We are developing a generative art application for casual creation on handheld devices, where user enjoyment is prioritised over the quality and/or utility of the abstract images they make. In an attempt to increase enjoyment, we have enabled the app to generate image titles, and we describe here the approach we take, along with details of some experiments we undertook to optimise it for efficiency and variety. The approach employs words relating to a machine vision classification and colour breakdown of the images, wrapped in the language of International Art English. When there is no suitable input from the machine vision analysis of an image, the approach generates text which is not necessarily related to the image. Such texts have been described linguistically as *pseudo-profound bullshit* statements, in work on the psychology of human judgement that we survey. We evaluate the title generation approach with a curation analysis, and end with a discussion of some potential benefits of employing bullshit in a Computational Creativity context.

1 Introduction and Motivation

Generative art [13, 16] has many uses, including professional art production, pushing forward the autonomy of Computational Creativity systems and supporting the creativity of users. Often overlooked is the fact that making art with a generative system can be a fun and entertaining pastime, over and above the value/utility of the pieces which are produced. In [3] and [4], Compton et al. have driven forward the study of *casual creators*, i.e., creativity support tools designed primarily with fun in mind. They describe casual creators as:

> ... interactive system[s] which encourage fast, confident, and pleasurable exploration of a possibility space, resulting in the creation or discovery of surprising new artefacts that bring feelings of pride, ownership, and creativity to the users that make them. [4]

As described in [1], we are building a casual creator app for the iOS platform called *Art Done Quick*. With this app, users are able to quickly and easily make, edit and share abstract art pieces produced in a generative fashion. This has been designed with a *fun-first* methodology, wherein we constantly prioritise the ease of use and fun aspects of the app over other aspects such as the sophistication of the images produced and editing power. There are two main modes of interaction with the app, namely on a large sheet where randomly generated images can be instantly added and viewed, and an editing screen where numerous changes can be made to a particular image.

[1] SensiLab, Faculty of IT, Monash University, Australia
[2] Game AI Group, EECS, Queen Mary University of London, UK
[3] Independent art historian and curator. www.blancapferrer.com

In the sheet mode, double tapping an image produces 8 variations, which are generated by mutating the underlying chromosome representation. Screenshots depicting the two modes are given in fig. 1.

Technical details of Art Done Quick are given in [1], and it is beyond the scope of this paper to describe the image generation or editing capabilities in detail. It is pertinent to know that the app employs a mobile version of the Resnet artificial neural network [8] for machine vision classification tasks. With this, it is occasionally able to project an image category such as *jellyfish* or *oscilloscope* onto one of the images produced, with high confidence (≥ 0.8). As the images are abstract, it can often be a fun exercise for users to determine whether they can see what Resnet sees in the image, and there is usually a moment of clarity when this happens. For instance, it takes some time to realise that the image in figure 2 – which Resnet categorises as depicting an *acoustic guitar* – can be viewed as the soundhole in a guitar, but once this is realised, it cannot be *unseen*.

We describe here an addition to Art Done Quick's functionality whereby it can expand a categorisation provided by Resnet into a title for the image. To do this, we have employed the language of International Art English to wrap suitable text around the category name. As the projection of a Resnet image category is a fairly rare occurrence, we have implemented the ability to produce a colour breakdown of an image, which opens up more possible title constructions. For images where there are no dominant colours and no suitable Resnet categorisation, images are given titles without reference to their content. The generation of these image titles has been influenced by the study of pseudo-profound bullshit, which – along with International Art English – is described in the next section.

We present details of our first approach to pseudo-profound pseudo-bullshit image title generation in section 3, including details of how we improved the efficiency of the colour breakdown process. Using a curation analysis, we evaluate the output of the title generation system in section 4, and identify limitations of the approach, while observing that it has the potential to be the basis for a more sophisticated generative method. We conclude in section 5 with a speculative discussion of how bullshit generation might be productively employed in casual creators and in Computational Creativity research and practice more generally. In particular, we argue that generative systems like Twitterbots [19] benefit from producing nonsensical (possibly bullshit), as well as interesting, output. We also argue that pseudo-profound titles – whether bullshit or not – can add to the value of abstract art pieces, especially if the pair is seen as a diptych with the title providing as much a platform for interpretation and/or aesthetic projection as the artwork. We end by describing planned improvements to the title generation approach in Art Done Quick, in the hope that it will become a much-loved aspect of the app after the public release planned for mid-2020.
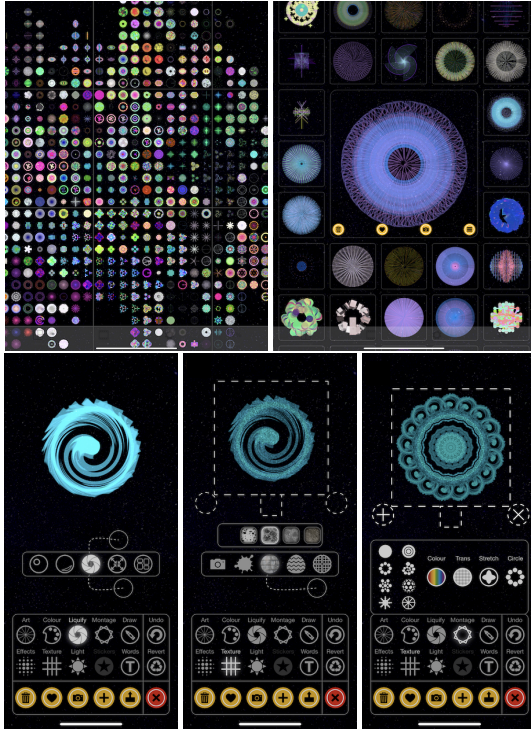
Figure 1. Screenshots from the sheet mode (top, iPad) and edit mode (bottom, iPhone) of the Art Done Quick casual creator.

Figure 2. Abstract art image categorised by Resnet as an acoustic guitar.

2 Background on Bullshit

In a controversial essay [14], Rule and Levine propose that people in the art world often use a special language they call *International Art English* (IAE), which "... has everything to do with English, but is emphatically not English". Aspects of IAE include: a tendency to use more rather than fewer words, to the point of redundancy; employing words such as *dialectic* for aesthetic rather than communicative purposes; and to assert authority and exclude the common reader from the elitist group of people that "get it". They provide support for this proposal with a linguistic analysis of 14,000+ press releases from the e-flux.com art subscription service, performed by the Sketch Engine [9] at sketchengine.eu. They highlight the relatively high usage of certain words such as 'space', 'biopolitical', 'transversal', 'autonomy', etc., in IAE, compared to the British National Corpus. As an example, an article currently in the e-flux journal is entitled: "Suspended Munition: Mereology, Morphology, and the Mammary Biopolitics of Transmission in Simone Leigh's Trophallaxis". Rule and Levine suggest a genealogy for IAE, originating (in part) from the translation into English of French poststructuralist texts.

In separate work [12], Pennycook et al. study the notion of *pseudo-profound bullshit* (PPB) as statements which may "seem to convey some sort of potentially profound meaning, [but are] merely a collection of buzzwords put together randomly in a sentence that retains syntactic structure." They provide as examples the following vacuous phrases: 'Hidden meaning transforms unparalleled abstract beauty' and 'Attention and intention are the mechanics of manifestation', with the latter being tweeted by a famous author who is often accused of bullshitting. Note that, as highlighted in a definition given by Frankfurt [7], bullshit is distinguished from lying because there is no deliberate subversion of the truth, although bullshitters often do so in order to hide ignorance or present authority when
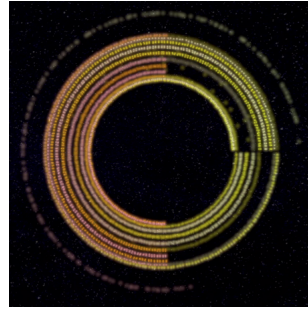
there is none. Using a bullshit receptivity scale, in a series of studies, Pennycook et al. showed that the propensity of people to judge bullshit statements as actually profound was associated with certain background experiences and personality traits. In particular (amongst other properties), people more receptive to bullshit are less reflective, lower in cognitive ability, more prone to conspiratorial ideation and are more likely to hold religious beliefs and endorse complementary/alternative medicine. In similar experiments, Erlandsson et al. show that bullshit receptivity and profoundness receptivity are positively correlated with each other, but profoundness-receptivity has a positive association, whereas bullshit-receptivity has a negative association with two types of prosocial behavior [6].

In [18], Turpin et al. attach pseudo-profound bullshit titles (such as *The Deaf Echo*) to computer-generated abstract art pieces, with the titles generated with no regard to the images they were attached to. The authors showed that the perceived profoundness of the images was increased by such titles, which was not the case when more mundane titles (such as *Canvas 8*) were given. The authors show that the effect transfers to abstract artworks created by human artists, and they further compare pseudo-profound bullshit with International Art English, reporting a large correlation for the profoundness ratings of images given PPB and IAE titles. They suggest that people use similar underlying cognitive mechanisms when interpreting PPB and IAE titles for artworks. Turpin et al. used an online system (at: noemata.net/pa/titlegen) for generating the art titles independently of the images. The details of the title generation system are not given either in [18] or the web pages, and it is not clear whether the method explicitly uses International Art English or not.

Leder et al. [10] investigated the influence of descriptive and elaborative titles on the appreciation and understanding of pairs of similar looking paintings by the same artist. Descriptive titles summarised the painting describing the depicted scene or pattern, e.g. *Dark zigzag lines on subdued background*, whereas elaborative titles supplied a possible interpretation, e.g. *Tears* or *Breaking into the technical era*. In an experiment of long exposure (90s), elaborative titles increased the understanding, but not the positive appreciation of the artworks by the participants. A second study further suggested that time is an important aspect in aesthetic appreciation: descriptive titles increased understanding more than elaborative ones in short exposures (1s), but for medium exposures (10s), it was vice-versa.

To the best of our knowledge, the generation of bullshit has not been the explicit goal of any system presented in Computational Creativity research circles. Of course, most generative systems producing text make nonsensical sentences at one stage or another, and some of this may fall into the category of bullshit. Sadly, this is often passed off as *surrealist* without any reference to the aims, concepts or methodologies of surrealism. In other contexts, generated gibberish is often passed off as poetry or some other language form, again with little regard to the literary culture being targeted.

3 Pseudo-Profound Pseudo-Bullshit Generation

A subset of the corpus of e-flux press releases compiled by Rule and Levine for their analysis of International Art English is pre-loaded into the Sketch Engine system that they used. The corpus is composed of around 5 million words from 170,000 sentences in 9,500 documents, and is tagged with the Penn Treebank tagset v2.5 [11]. Sketch Engine is able to extract a set of keywords from the e-flux corpus, and we chose a subset of 50 keywords to be potentially useful in pseudo-profound bullshit art titles which employ International Art English (IAE). The first set of these keywords was identified in [14] as being particularly over-used in IAE, e.g., abstract, dialectical, radically, space, manifestation, transversal, etc. The second set was chosen to be generally high-brow terms such as intellectual, contemporary, expression, proposition, autonomy, etc. The third set were general art terms: juxtaposition, paint, draw, render, portrait, etc., and the final set were abstract notions such as pain, obsession, void, imagine, etc. The full set of keywords is given in the appendix.

Given a target word, Sketch Engine can be used to perform a number of linguistic analyses of the e-flux corpus relative to the word. In particular, it can extract a thesaurus of words which are correlated with a given word W along with a confidence score between 0 and 1 for each extracted word. For instance, the first three thesaurus terms for 'abstract' are 'sculptural', 'conceptual' and 'narrative'. Note that this contrasts with the thesaurus words for 'abstract' given by analysis of the English Web 2015 (enTenTen15) corpus, the first three of which are 'visual', 'experimental' and 'original'. Sketch Engine can also provide a *word sketch* for W, which contains words which are associated with W in the corpus. For instance, the word sketch for 'abstract' contains a set of *modifiers* which can appear directly before it, such as 'resolutely', 'completely' and 'purely'. For each of the 50 keywords, we used Sketch Engine to extract both a thesaurus and a word sketch, combined the results and added them as resources to Art Done Quick.

Using the keywords, we compiled ten templates which can generate text to wrap around a word describing an image categorisation that Resnet ascribes to an image with high confidence. These templates have been kept relatively simple, in order to avoid obfuscating too much the image category that Resnet finds, so that users can experience the puzzle of trying to see the image in the same way that Resnet does. As examples, three of the templates are:

- untitled (#object)
- #object abstraction
- the autonomy of the #object

To generate a title from a template, given the label C of a Resnet classification, Art Done Quick substitutes C for the placeholder #object and a randomly chosen word from the thesaurus for any keyword in the template. As an example, the third template above gets expanded to 'The freedom of the jellyfish' for the second image in the first row of figure 6, because 'freedom' is a thesaurus word for 'autonomy'.

There are a number of limitations to generating titles in this fashion, including: (a) Resnet only finds a high-confidence classification for around 1 in 10 randomly generated images (b) given the constrained nature of the images, the classification is often repeated, for instance oscilloscope and analog clock are often projected onto images, and (c) the trained model for Resnet that Apple make available for use in apps has been trained on the competition version of ImageNet [5], which has 1,000 categories of images chosen to help differentiate machine vision systems rather than being of general use. To highlight point (c), we note that Resnet is able to identify more

than 200 animals, including 22 different types of terrier (dog) – the utility of this in our context is quite limited.

Given these limitations, we used a colour breakdown to provide further machine vision detail for the title templates. To do this, a sample of 1,000 pixels is taken from the (500×500 pixel) image and a list of named colours that the pixels are closest to is compiled. The possible colours are taken from htmlcsscolor.com and include 1,639 named colours, many of which have a lyrical nature (such as lavender rose), others have artistic connotations (such as moody blue) and others represent objects such as rhino and eagle. The nature of the images produced by Art Done Quick means that a majority of pixels will be black. After some initial experimentation, we determined that if 50 pixels or more from the 1,000 pixel sample were mapped to the same closest non-black colour, then the colour would be clearly visible in the image, and the colour name could be used fairly safely in the image title. To harness this colour analysis, we added another 10 templates which use the name of the primary colour in the same way as the Resnet classification. We also constructed 10 templates using the two most prevalent colours in the image.

For many images, there are too many colours with too few pixels for any one of them to be particularly dominant, and likewise no high-confidence Resnet classification. In these cases, Art Done Quick generates a title as a pseudo-profound bullshit statement without reference to the image. To enable this, for each of the 50 keywords, we constructed a template which embeds that keyword and possibly others. For instance, the template for the keyword 'abstract' is 'an abstract study in autonomy', and when expanded with thesaurus words for all three keywords in the template, yields titles such as: 'A poetic development in morality' and 'A unique culture in solitude'. For this template, there are 204,424 different instantiations possible, given the thesaurus terms from Sketch Engine.

We extended the template mechanism slightly to enable information from the word sketch for a keyword to be employed. In particular, if the keyword 'modifier' appears in a template, then it will be substituted by a modifier from the word sketch for the word after it in the template. We also enabled the templates to specify options such as they/we/I/you to be chosen randomly when the template is expanded. The templates for the pseudo-profound bullshit titles were derived under consultation with an art historian (third author), and are given in full in the appendix. They were derived in part by surveying the titles given to abstract artworks over the last 50 years, and generalising any patterns found.

3.1 Improving Efficiency

Initial experiments with the title generation highlighted that the process took nearly 2 seconds on an iPhone XS (a target device for Art Done Quick), and that this waiting time lowered the enjoyment gained from being given the titles. On investigation, we found that the time for the application of Resnet and the templating system were both negligible, but the colour decomposition of the image took on average 1.8 seconds. This is because the process checks each of 1,000 pixels against each of 1,639 colours, hence for each image, 1.6m Euclidean distances (in RGB space) were being calculated.

To increase efficiency, we first implemented a hashing system whereby a pixel's colour, C, in RGB form, is first mapped to the closest, K, of a small number (≤ 100) of key-colours. K is associated with another small set of colours, which are searched over for the closest match, M, to C. While this reduces the accuracy of the assignment of named colours C, it greatly increases the efficiency. The accuracy of the colour breakdown of an image is not critical,
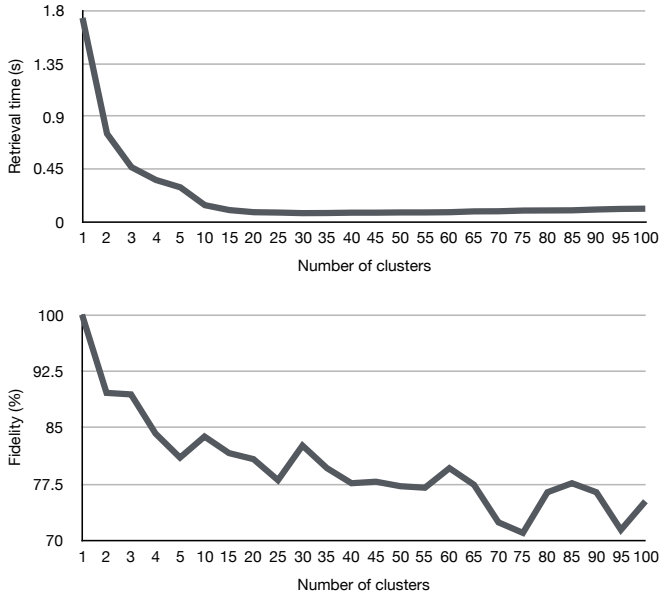
Figure 3. (a) Change in retrieval time as the number of clusters increases. (b) Change in fidelity when compared against 1-cluster setup, as the number of clusters increases.

Figure 4. Clustering of the named colours into 23 clusters, with cluster size at the top and key-colour for each cluster at the bottom.

i.e., users are unlikely to be able to tell the difference between honey and caramel or lavender pink and lavender rose, and would likely not notice an imperfect assignment of a colour to an image.

To generate the set of key-colours for the hash-table, we used k-means clustering over the set of 1,639 colours and experimented with different values of k from 1 to 100. In this way, we were able to use the centroids of the clusters as the key-colours and the members of each cluster as the set of colours associated with the key-colour (centroid). To see the improvement in efficiency, we created a clustering for each of k from 1 to 100 and used the key-colour lookup approach to determine the colour of 500 pixels in each case. Averaging over the time taken, the results are given in figure 3(a), and we see that the 1.8 seconds taken with only one cluster (i.e. all the colours) reduces to around 0.05 seconds when k is 20 or more. We also tested the change in *fidelity* as k increased, by recording the percentage of times that the most-prevalent colour determined by the hashing procedure was the same as that determined by testing against each of the 1,639 colours (the try-all approach). The results are presented in figure 3(b), and we see that between $k = 20$ and and $k = 50$, the more efficient approach projects the same most-prevalent colour as the try-all approach 77% of the time or more.

Note that k-means clustering is stochastic in nature, so re-running the clustering process for the same value of k might produce a better clustering, in terms of fidelity. Hence, to improve accuracy while keeping the efficiency gains, we searched for a (near) optimal clustering with k between 20 and 50 by generating 20 different clusterings for each k and recording the one with the highest fidelity. In this way, we found a clustering of the 1,639 colours into 23 clusters, with a fidelity of 84.6%, which we felt was high enough, given the lack of concern over accuracy in colour decomposition. The clustering was extracted and made available as a resource to Art Done Quick to load, rather than generating it, at the start of a session. A visualisation of the clustering is given in figure 4 and we see that the largest cluster is
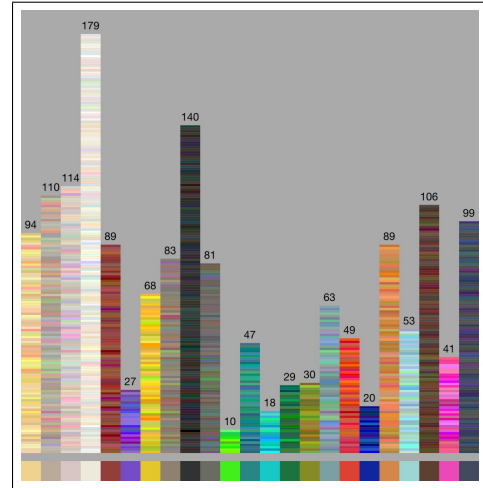
of white/light grey colours, with 179 named colours in it, while the smallest cluster, of bright green colours, has only 10. This means that the worst case scenario is the calculation of 23 + 179 = 202 Euclidean distances for a pixel, and the average is much lower.

To further reduce the time taken to break down an image into a set of named colours, we investigated reducing the sample size from 1,000 pixels per image. Ranging from 100 to 900 pixels sampled, we tested the fidelity of the hashing method described above when compared to the method with 1,000 pixels sampled. We averaged the fidelity and the execution time (on an iPhone XS) over 250 images where there was at least one dominant colour (i.e., with 50 or more pixels – out of 1,000 randomly sampled pixels – mapped to the same named colour). Fidelity was recorded as the percentage of times when the method with a smaller sample size chose the same most-prevalent colour as the 1,000 sample method. The results are given in figure 5. Using these results, we chose a sample size of 300, as this achieves the same colour mapping 86.25% of the time, while being three times as fast as the 1,000 sample method. Overall, we reduced the processing time for an image from an average of 1.8 seconds to 0.023 seconds, while still maintaining an acceptable quality level for the colour breakdown.
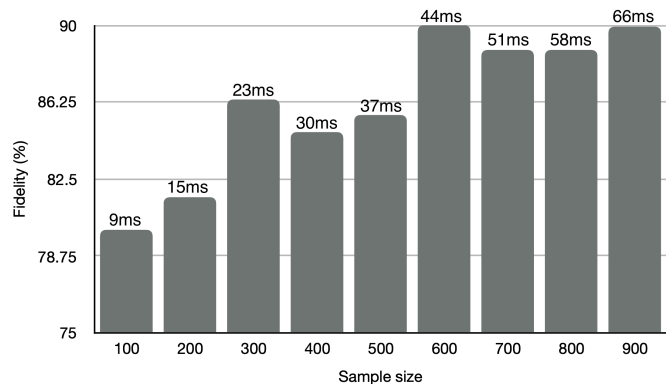

Figure 5. Fidelity and speed (given in milliseconds above the fidelity bar) of the colour analysis method, as sample size increases.

4 Evaluation

The title generation approach presented here is not particularly sophisticated yet, and the output from the system is not ready for testing against users. Hence we decided that a straightforward *curation analysis* was appropriate at this stage. In [2], Colton and Wiggins suggest that the output of a creative system can be used to calculate a *curation coefficient* which counts the proportion of artefacts produced which satisfy a binary subjective judgement (usually by the system's author) of value. This is a sensible measure for us, and these results will act as a benchmark for further curation analyses of improved image title generation methods in Art Done Quick, before we expose the results to users for a more thorough evaluation.

In [18], Turpin et al. use pseudo-profound bullshit titles for generative art pieces in their psychology studies, with the titles generated by an online system and selected ones judged by the authors as being pseudo-profound bullshit or not. However, they do not provide a reliable way of telling whether a particular phrase is pseudo-profound bullshit or not, and give no details of how their bullshit was generated. We found that many titles generated by our approach were *divisive*, i.e., one person would think they were bullshit, (a grammatically correct meaningless collection of buzzwords), while another would be able to interpret some meaning in them.

Given this subjectivity, we employed two evaluators (first and second author) to assess the same 100 titles independently. Both evaluators were instructed to categorise image titles into (a) grammatically incorrect, hence unreadable, (b) bullshit as above, or (c) one of any number of additional categories that they felt the titles should be put into. In a preceding exploration, the first evaluator looked at around 200 images with titles, and chose to categorise them into four types: grammatically incorrect, bullshit, conceptually interpretable and emotionally interpretable. Example images/titles in these categories are given in figure 6, along with images/titles where Resnet confidently classified the image, and images with titles using the colour breakdown. The following are examples:

- The title 'To endlessly design paradox' was assessed as being conceptually interpretable (and hence not bullshit), as it brought to mind a deity setting puzzles for humanity.

- The title 'Painting of the tenth emptiness' was read as melancholy due to the emptiness reference and implication of many such emptinesses, and was assessed as being emotionally interpretable.

- The title 'The sensibility of subjectivity' was deemed as too vacuous for meaningful interpretation, and was hence assessed as bullshit.

The second evaluator assessed the same 100 image titles without reference to their corresponding images, and placed the titles into one of four categories: ungrammatical, bullshit, descriptive and poetic. A title was assessed as descriptive if it was a literal description of what the image depicts or makes a reference to a formal aspect of creation, like a technique or material, in an unpretentious way. A title was assessed as poetic if it exhibited the right amount of depth and interpretability. The second evaluator noted that – especially for titles that use neither Resnet label nor colour name – the line between 'poetic' and 'bullshit' is very thin. The decisive difference between bullshit and descriptive or poetic titles often lay not in what was said exactly, but how it was phrased. It is possible, the evaluator noted, to turn a bullshit title into a 'descriptive' one, e.g., rephrasing 'The restructuring, the renowned scenery' to 'Restructuring renowned scenery' re-establishes the title's purpose. Rather than obfuscating, it now sheds light onto an internal reflection on formal artistic expression and in-

tention which is communicated through a work of art.

The results of the evaluations are given in table 1(a) for evaluator 1 and table 1(b) for evaluator 2. The rows in each table separate the titles into three subsets: those produced in reference to an object identified in the image by Resnet (of which there were 13 in the 100), those referring to the colours in the image (32), and those not referring to the image at all (55). This makes the object ones rather special, but unfortunately this effect is ruined by the repetitions, with sea urchin seen 3 times and coil seen 4 times. In [1], we describe ways to increase the yield and the variety of Resnet-labelled images.

As expected, the reviewers assessment of pseudo-profound bullshit differed considerably, with the first reviewer categorising 30% as bullshit, compared to the second reviewer's 53%. In total, 62% of the titles were judged by evaluator 1 as being in the conceptual/emotional sweet spot, but only 38% of titles were judged to be in the descriptive/poetic sweet spot by evaluator 2. It seems clear that removing ungrammatical titles entirely from those produced by Art Done Quick is an imperative for a baseline version. However, as argued below, it's not clear that reducing bullshit titles – as much as possible given the subjectivity in determining what is bullshit – is desirable. Certainly, we will be trying to make non-bullshit titles more prevalent, and we describe some possible improvements to this end in the next section. One area for improvement will be in producing emotional titles (as per evaluator 1) and poetic titles (as per evaluator 2), as these were assessed as being the most valuable kinds of titles, yet were fairly uncommon (21% and 8% respectively). To increase the yields for these types of titles, it may be sensible to view emotional and poetic as different, given that there were only three titles which overlapped both, namely: 'A ballet of ancient fear', 'The wasteland, the strong residue' and 'The impact of raw umber'.

In overview, the curation analyses show that certain titles were generated which work very well as accompaniments to abstract art images, hence we believe the template approach using International Art English shows much promise as a basis for a more sophisticated approach, and we describe below some improvements we plan.

5 Conclusions and Future Work

While the studies in [12] and [18] make direct reference to International Art English (IAE), and investigate what they term 'bullshit' based on Frankfurt's conception, we understand this notion to be broader. IAE is related to bullshit in the sense that it is used to obfuscate rather than explain, in order to make something appear more impressive than it is. This is not the goal for the titles we generate. Instead, we aim for thought-provoking titles, that ideally provide a new perspective on a depicted scene or object, or an abstract image.

Taken overall, the rudimentary image title generation in Art Done Quick produces fully bullshit texts only some of the time. Given the usage of IAE, the majority of the titles are pseudo-profound and whenever they do not make explicit reference to the depicted image through the Resnet label, they adhere to Frankfurt's conception of retaining no regard for the truth. Hence we use *pseudo-profound pseudo-bullshit* to describe the types of titles produced. It would be easy to dismiss bullshit as being something to be avoided at all costs in natural language generation, as such statements are, by definition, devoid of meaning. Moreover, the term bullshit is commonly used as a derogatory term in general usage, and often lies are called out as bullshit. However, the situation is not as black and white as it may seem. The following are reasons that we might want to keep a certain amount of pseudo-profound bullshit in Art Done Quick and other Computational Creativity systems.

- Visual art doesn't have to be representative to be of value, as abstract art provides a platform for interpretation which is much valued. When a title is attached to an image, it is normally done so in a complimentary fashion, i.e., to help in understanding the image. This doesn't have to be the case, though. When an abstract title is attached to an abstract image with no regard to it, it seems sensible to think of the pair as a diptych of equally valuable opportunities for interpretation, rather than the title as being subordinate to the image. If the title is bullshit, then finding an interpretation may not be easy/possible, but the text could still offer something lyrical or aesthetically pleasing in other ways, as is suggested is true for aspects of the text in art press releases [18].

- In Art Done Quick and similar systems, a user may be shown dozens of images with titles in fairly rapid succession. This is analogous to *twitterbots* [19], which produce endless numbers of often nonsensical short texts. Rather than the nonsensical ones being pointless, they provide a context in which a reader may feel ownership if they find a gem hidden between low quality outputs. We can hypothesise that for a bot to be successful, there has to be as much wheat as there is chaff (to use an English idiom) in its output. In particular, the fact that the texts are generated by computer means that there is a good chance that the twitterbot's programmer has not seen a particular tweet, so a keen twitter follower could be the first to see a gem posted by a twitterbot, and there is value in being the first to retweet such a post. In the context of Art Done Quick and other casual creators, the text generation is done on the user's device, so they are assured of being the first to see a particularly good image title, and we believe there needs to be a sufficiently high proportion of bullshit titles generated for the full benefit of gem ownership to be realised.

- As we have seen in the curation analysis above, one person's bullshit title could be another person's profound statement. Being too strict in avoiding bullshit titles could be a mistake in this context. Moreover, even if an image title is universally assessed as bullshit, some people may find humour in the utter vacuousness of the text. This may be as a tongue-in-cheek critique of the art world from which International Art English stems, or of social media influencers, politicians and others who regularly bullshit, or of AI developers and their inability to understand the culture to which they apply their work. In a casual creator app, where fun and entertainment are prioritised, such humour would be well placed.

Before testing whether the above reasons for Art Done Quick to generate bullshit are valid, we plan a number of improvements to its title generation method. In particular, some hand tweaking of the word sketches and thesaurus entries, additional tagging of words, and improved natural language processing should reduce the production of ungrammatical titles, possibly removing them entirely. Moreover, even though everything has to be done quickly in Art Done Quick, there is time to produce multiple titles for a particular image from which to choose the best one according to a fitness function, which may include some reference to profundity and bullshit. We further plan to use natural language resources such as a pronunciation dictionary and techniques such as sentiment analysis to make the titles more lyrical and poetic, whether they are bullshit or not. There is a wealth of text generation approaches that we plan to learn from in order to implement better generative techniques, and we will look in particular areas such as slogan generation [17]. We also plan to investigate machine vision models other than Resnet, which can identify more visual aspects of an image, and image captioning models [15].

In her writings on Casual Creators [4], Compton has stated that the most successful apps often have magic-like elements which surprise users, either through enhanced control, the production of unexpected output or something powerful in the system. Often the generative aspects of casual creators provide the most obvious magical moments, and in [1], we describe how the very fast generation of images (so that a set of quite varied abstract images can be instantly added to the sheet in Art Done Quick), and the mutation of images into variations on a theme may fit the bill for such magical aspects. We hope that the generation of image titles will be the third magical element of Art Done Quick, and we expect that this will be because of, rather than in spite of the bullshit generated by the system.

Acknowledgements

REFERENCES

[1] S Colton, J McCormack, S Berns, and E Petrovskaya, 'Adapting and enhancing evolutionary art for casual creation', in *Proceedings of the Ninth International Conference on Artificial Intelligence in Music, Sound, Art and Design*, (2020).

[2] S Colton and G Wiggins, 'Computational Creativity: The final frontier?', in *Proceedings of the 20th European Conference on Artificial Intelligence*, (2012).

[3] K Compton, *Casual Creators: AI Supported Creativity for Casual Users*, Ph.D. dissertation, University of California, Santa Cruz, 2019.

[4] K Compton and M Mateas, 'Casual creators', in *Proceedings of the 6th International Conference on Computational Creativity*, (2015).

[5] J Deng, W Dong, R Socher, L-J Li, K Li, and F-F Li, 'ImageNet: A large-scale hierarchical image database', in *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, (2009).

[6] A Erlandsson, A Nilsson, G Tinghog, and D Vastfjall, 'Bullshit-sensitivity predicts prosocial behavior', *PLOS One*, **13(7)**, (2018).

[7] H Frankfurt, *On Bullshit*, Cambridge University Press, 2005.

[8] K He, X Zhang, S Ren, and J Sun, 'Deep residual learning for image recognition', in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, (2016).

[9] A Kilgariff, V Baisa, J Busta, M Jakubicek, V Kovar, J Michelfeit, P Rychly, and V Suchomel, 'The Sketch Engine: Ten years on', *Lexicography*, **1(1)**, (2014).

[10] H Leder, C-C Carbon, and A-L Ripsas, 'Entitling art: Influence of title information on understanding and appreciation of paintings', *Acta psychologica*, **121**(2), (2006).

[11] M Marcus, B Santorini, and M Marcinkiewicz, 'Building a large annotated corpus of English: The Penn Treebank', *Computational Linguistics*, **19(2)**, (1993).

[12] G Pennycook, J Cheyne, N Barr, D Koehler, and J Fugelsang, 'On the reception and detection of pseudo-profound bullshit', *Judgment and Decision Making*, **10(6)**, (2015).

[13] *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*, eds., J Romero and P Machado, Springer, 2008.

[14] A Rule and D Levine, *International Art English*, Triple Canopy, 2012.

[15] P Sharma, N Ding, S Goodman, and R Soricut, 'Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning', in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, (2018).

[16] H Takagi, 'Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation', *Proceedings of the IEEE*, **89**, (2001).

[17] P Tomasic, M Znidarsic, and G Papa, 'Implementation of a slogan generator', in *Proceedings of the Fifth International Conference on Computational Creativity*, (2014).

[18] M Turpin, A Walker, M Kara-Yakoubian, N Gabert, J Fugelsang, and J Stolz, 'Bullshit makes the art grow profounder', *Judgment and Decision Making*, **14(6)**, (2019).

[19] T Veale and M Cook, *Twitterbots: Making Machines that Make Meaning*, MIT Press, 2018.

Appendix: Templates with Keywords in Bold

an **abstract** study in autonomy
a balanced **abstraction** in space
draw, see, **analyse**
the modifier **artist**
the proposition of **autonomy**
balanced and transversal modifier obsession
colorful tension, dialectical portrait
conceptual yet formal transversal tension
the artist's **contradiction**
modifier **contemporary** portrait
critical expression of the slogan
modifier **dialectic** and modifier **dialectic**
the **dialectical** impression

they/we/I/you **discover** the invention
to **discuss** and emerge
draw, not paint
they/we/I/you **emerge** as a modifier hierarchy
the **expression** of obsession
increasingly **feminine** memory
they/we/I/you **forget** the slogan
they/we/I/you see the **formal** expression
study in formal **hierarchy**
to **imagine** a modern space
manifestation of an **impression**
increasingly modern and reflexive

abstract **invention** of modifier reflection
an **intellectual** juxtaposition
painful modifier **juxtaposition**
modifier **landscape** piece
manifest reflection of the void
the **manifestation** of memory
a **memory** of modifier autonomy
modern portrait
an **obsession** to render
discover the **painful** expression
to see and **paint**
radically render the piece
reflexive portrait in modifier space

the **proposition** of abstraction
to **radically** see tension
a **reflection** in dialectical space
they/we/I/you **render** to see
they/we/I/you **see** the tension
simultaneously paint and draw
a **slogan** of modern obsession
study in **space**
study of the transversal void
increasingly the **tension** manifests
abstract and **transversal** piece
the **void**, the contemporary **void**
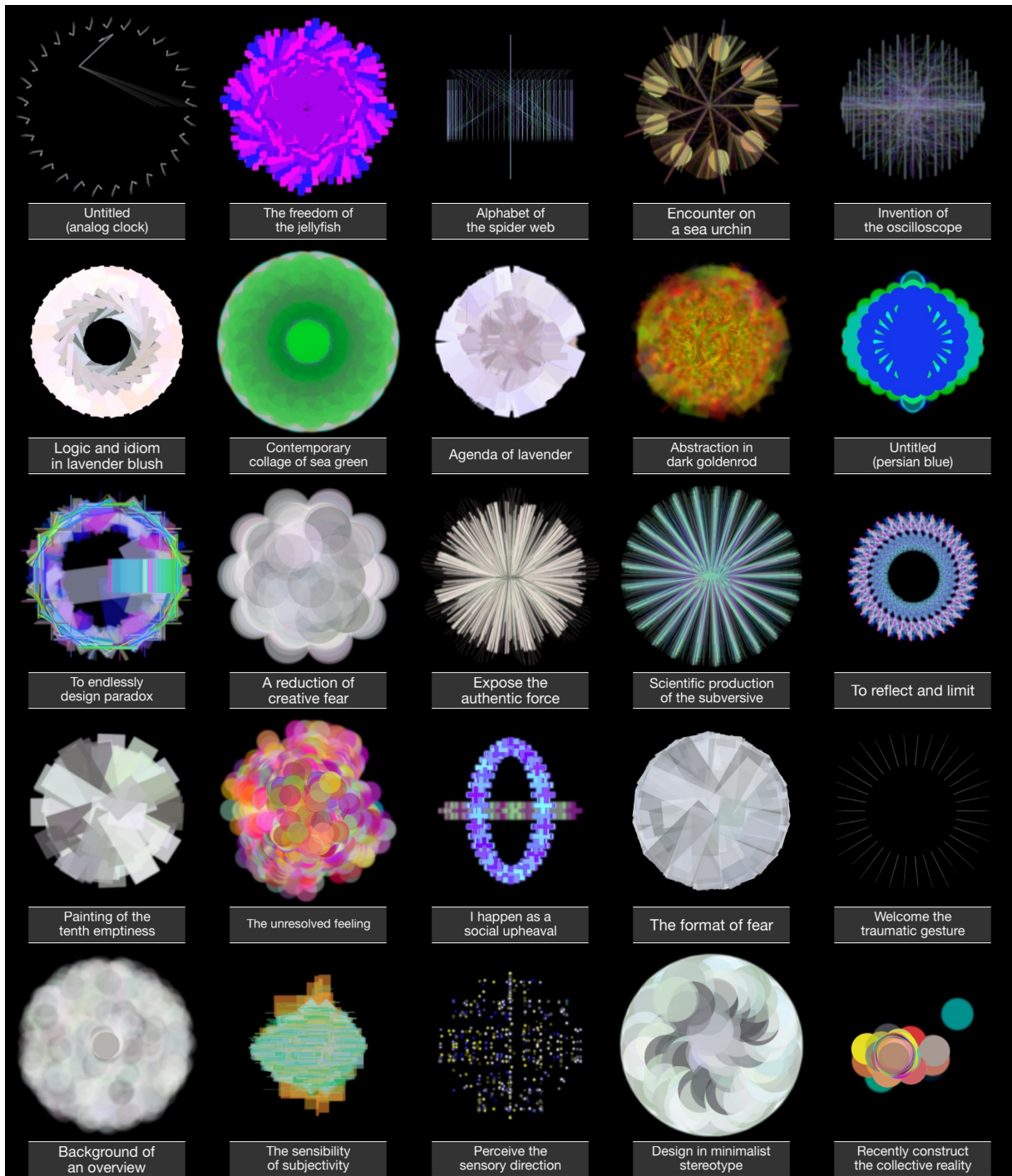
Figure 6. Example generative art images, along with the titles produced by the IAE template approach. First row: titles using the Resnet image categorisation. Second row: titles using the primary colour in the image. Third row: IAE titles not related to the image, but interpretable conceptually. Fourth row: IAE titles not related to the image, but interpretable emotionally. Fifth row: uninterpretable IAE titles not related to the image (pseudo-profound bullshit).

13

	Ungrammatical (8)	Emotional (21)	Conceptual (41)	Bullshit (30)
Object (13)	Artistic jellyfish in element Truly correspond sea urchin Unique digital clock occur in area	Dynamic sea urchin free in process Geographic birdhouse structure	Development of a bubble Sea urchin representation Untitled (coil) Untitled (coil) Model of a soccer ball Activity of a coil	Rather determinate diaper Potentiality of the coil
Colour (32)	Fiction in nebula Unfamiliar and costume light grey Life of east side	The impact of raw umber Research in morning glory Free speech red We understand the gunpowder Self and spectre in tower grey The spiritual difference of light grey	Medium slate blue Constructivist and transversal grey Granny smith apple Reflection of dodger blue Untitled (manatee and topaz) Approach in blue violet Expectation of nero Collection in whisper Conversation in city Version of lavender blush Transatlantic and animate hit grey Medium red violet Studio architecture of turkish rose Documentary of bombay	Potentiality and articulation in seal brown Productive and ceramic blackcurrant Documentary space of deep magenta Global orchid, innovative plum Trope and quest in malachite Social poster of panache Axis and topic in ghost Negotiation of mamba Excellence and thinking in concrete
None (55)	An understanding in diplomatic theme Long architectural and reflexive	A seduction to utilize An overtone of creative obsession Corporeal careless manner A powerful manner The wasteland, the strong residue A happiness to characterise Again ominous representation Survey of the comfortable donors I abandon the imaginaries The character of happiness The ecological interaction of summer sky A ballet of ancient fear We hear the upheaval	The legacy of drive A historical examination To propose and imagine Give, choose, mirror Monumental society location A digital book in morality Dialogue in technical code Exhibit, establish, scrutinise Unique process area They include the transformation To propose and present Generic extent, queer representation You think the combination Connect, not succeed The programme's problem A direct accumulation They inhabit the storytelling Yet indeterminate material The visual programme Artist contribution of storm dust Experiment in radical millieu	About collective projection The restructuring, the renowned scenery About such material Foremost present drawing Own spirituality and playful consumerism Very aesthetic interview I change the complexity Residency in psychological limit Collective painting of highball The emptiness, the own renewal We base as a cultural stereotype Performance in presentation Bodily same syntax Intervention of the costume homeland Then cultural and reflexive An impact in transatlantic structure Collage in current felinity Mode of a visibility Imply collection of the scenery

	Ungrammatical (9)	Descriptive (30)	Poetic (8)	Bullshit (53)
Object (12)	Truly correspond sea urchin Unique digital clock occur in area	Artistic jellyfish in element Dynamic sea urchin free in process Sea urchin representation Untitled (coil) Untitled (coil) Model of a soccer ball Activity of a coil Rather determinate diaper	Development of a bubble	Geographic birdhouse structure Potentiality of the coil
Colour (32)	Transatlantic and animate hit grey	Unfamiliar and costume light grey Life of east side Self and spectre in tower grey Medium slate blue Reflection of dodger blue Constructivist and transversal grey Granny smith apple Untitled (manatee and topaz) Approach in blue violet Conversation in city Version of lavender blush Medium red violet Studio architecture of turkish rose	Fiction in nebula The impact of raw umber Free speech red Documentary of bombay	Research in morning glory We understand the gunpowder The spiritual difference of light grey Expectation of nero Collection in whisper Potentiality and articulation in seal brown Productive and ceramic blackcurrant Documentary space of deep magenta Global orchid, innovative plum Trope and quest in malachite Social poster of panache Axis and topic in ghost Negotiation of mamba Excellence and thinking in concrete
None (56)	Long architectural and reflexive Again ominous representation You think the combination We base as a cultural stereotype Bodily same syntax Then cultural and reflexive	A historical examination Dialogue in technical code Exhibit, establish, scrutinise Yet indeterminate material Artist contribution of storm dust Experiment in radical millieu Very aesthetic interview Performance in presentation Collage in current felinity	An understanding in diplomatic theme The wasteland, the strong residue A ballet of ancient fear	A seduction to utilize An overtone of creative obsession Corporeal careless manner A powerful manner A happiness to characterise Survey of the comfortable donors I abandon the imaginaries The character of happiness The ecological interaction of summer sky We hear the upheaval The legacy of drive To propose and imagine Give, choose, mirror Monumental society location A digital book in morality Unique process area They include the transformation To propose and preset Generic extent, queer representation Connect, not succeed The programme's problem A direct accumulation They inhabit the storytelling The visual programme About collective projection The restructuring, the renowned scenery About such material Foremost present drawing Own spirituality and playful consumerism I change the complexity Residency in psychological limit Collective painting of highball The emptiness, the own renewal Intervention of the costume homeland An impact in transatlantic structure Mode of a visibility Imply collection of the scenery

Table 1. Curation analysis of 100 image titles, undertaken by two evaluators. Titles are broken down into evaluator-defined categories (columns) and content (rows). 'Object' indicates that a Resnet classification of the image was used and 'Colour' indicates that the names of the most prevalent colour(s) were used.

Darwinian Creativity as a Model for Computational Creativity

Alice C. Helliwell[1]

Abstract. This philosophical paper examines the Darwinian account of creativity as a model for assessing computational creativity. It will first establish a Darwinian account of creativity using Simonton's [1] model. It will then apply this model to popular image-producing AI, Generative Adversarial Networks, and the promising Creative Adversarial Network, both used in the computational production of 'artworks'. The paper will argue that these networks are compatible with a Darwinian account of creativity, due to the presence of blind variation within the networks, a key component of Simonton's model. The paper will then address some initial objections. The aim of this paper will ultimately be to assess whether the AI systems are compatible with the Darwinian model of creativity, and in the process explore Darwinian creativity as a potential standard for testing computational creativity.

1 INTRODUCTION

The Darwinian model of evolution is thought to have wide and varied applicability [1]. Following Campbell [2], Simonton [1] suggested the application of Darwinian theory to creativity, given the arguable creativity in the evolutionary process. This paper will examine the model of Darwinian creativity suggested by Simonton. It will then apply this model to two 'creative' image-making Artificial Intelligence systems: Generative Adversarial Networks (GANs) and Creative Adversarial Network (CAN) [3]. This paper will assess whether the AI systems are compatible with the Darwinian model of creativity. Initial objections to the use of this model will then be addressed, followed by an assessment of the Darwinian model of creativity as a potential tool for evaluating the creativity of computational systems.

2 DARWINISM AND DARWINIAN CREATIVITY

There are two types of Darwinism: the first has been developed in the purely biological sense, with a focus on genetics, molecular biology and behavioural science [1]; the second type of Darwinism provides, according to Simonton [1] a model which can be applied to many developmental processes. This includes processes which are not purely biological, such as knowledge acquisition. This model consists of blind variation, selection and retention. This second type of Darwinism has been applied as a framework to a variety of processes, such as Skinnerian operant conditioning and evolutionary epistemology [4]. This Darwinian model can also be applied to creativity.

Simonton [1] proposes that there may be a basis for a selectionist model of creativity. This model suggests that in the case of humans, there is a psychological mechanism for producing variation, either through recombination or mutation. The outcomes of this variation then go through a selection process; in evolution this would be through sexual selection. In other fields, such as creativity, this selection process would be through the outcome being assessed against necessary criteria. Finally, successful variations are retained in the system.

The variation component is a controversial element of the model [1]. In order to be Darwinian, variation must be "blind" to the selection criteria; it must be as likely to be successful as unsuccessful (non-teleological) [5]. Campbell [2] argued that this blind variation could be seen in creativity. This does not mean that variation must be random, rather that likelihood of success is random. Just as in biological variation, some combinations or mutations may be more likely than others to occur, but they will not necessarily produce better adaptations [1]. It is important to note that this blindness applies to the production of variation, not the selection of successful variations, which will not result in equal likelihood of success.

There is some evidence to suggest that this is how human creativity works. Sternberg and Davidson [6] suggest that random priming from environmental stimuli produces a blind variation effect in human creativity; the input is somewhat unrelated to the task and thus provides an element of blindness. Simonton [1] notes that this fits with a large amount of the anecdotal evidence from creative individuals regarding their creative process.

Simonton [1] also addresses the possibility that computer creativity could follow a Darwinian model. Boden [7] states that computational creativity does not typically follow a Darwinian understanding of creativity, rather it tends to use logical processes or heuristic principles. Boden [7] did state however that with the advent of parallel processing and connectionism this may become more possible. Since Boden wrote on this issue, these technologies have advanced considerably, however, much of computational creativity does not focus explicitly on following a Darwinian model.

Genetic programming is one form of Artificial Intelligence that follows an evolution-based model based on Mendelian genetics (a mathematical approach which forms the basis for our understanding of genetic traits, a step further than Darwinism - how genetic recombination occurs) [8]. Genetic algorithms can have mutations added in each generation, which are then tested against the programmed selection criterion. Mutations which produce the best results are fed into the next generation. This process continues until the best fitting genotype is found [9, 10]. This type of programme has been used to rediscover key scientific discoveries [1] however it has not been applied to artistic creation.

---

[1] Dept. History and Philosophy of Art, Univ. of Kent CT2 7UG, UK. Email: `ach53@kent.ac.uk`

According to Boden [7], in order to meet the criteria for a Darwinian computational process, there must be some method of blind variation present within the model. Boden suggests that to reach the high levels of creativity reached by humans, ideational variation must also be a factor [7]. Ideational variation in creativity refers to the variation of ideas, not merely variation within existing rules or constraints [1].

3 OBJECTIONS TO DARWINIAN CREATIVITY

Simonton [1] addresses four potential objections to the Darwinian model of creativity. The first is the idea that creativity rises from sociocultural state rather than from individuals; if one individual had not come up with the idea, someone else would have. Simonton states that this does not offer any threat to the Darwinian model [1].

A second objection is that the Darwinian model of creativity eliminates the role of individual volition; there is no space in the model for the will of people. Simonton [1] argues that the role of individual will does not eliminate the need for variation, as one cannot *will* a creative breakthrough to occur, blind or environmental variation is still needed to stimulate variation.

The third potential objection to Simonton's Darwinian creativity is that creativity can be simply explained by human rationality. Simonton [1] discusses that with increased complexity, rationality becomes less applicable to solution-finding. Blind variation and testing theory are still applicable, particularly in cases of extreme novelty and complexity.

Finally, Simonton [1] discusses an objection based on domain expertise; the idea that those who have expertise in a field no longer need trial and error. In this Simonton refers to the original nature of creativity [11, 12]. There must be a balance of originality and expertise in creativity, which still leaves room for variation and non-expert input. Simonton also suggests that creativity cannot be improved upon with expertise; one cannot get more creative with age or experience [1].

4 GENERATIVE ADVERSARIAL NETWORKS

In order to assess the possible application of a Darwinian model of creativity to AI, it is necessary to test this application. This paper will examine two image-production AIs: Generative Adversarial Networks and Creative Adversarial Networks. These particular systems will be examined as they offer a plausible case for artistic creativity in AI.

Generative Adversarial Networks (GANs) are a form of Artificial Intelligence which utilize machine learning to produce 'artistic' or 'photographic' images [13]. They consist of two parts: the generator and the discriminator. The discriminator is fed the training images: in this case, images of human artworks. The discriminator learns to distinguish things that fit into the model of "human artwork".

The generator does not have access to the training set, and is blind to the discriminator's rules about what is or is not an artwork. The generator initially begins producing random images, with randomness drawn from a noise vector. These are fed into the discriminator. The discriminator assesses the image in comparison to the model it has built based on the training set. The discriminator feeds back a score into the generator,

corresponding to whether it thinks the image is a "fake" artwork, or a real one. This score is used by the generator to adjust future outputs through adding weights to the algorithm, which increases the probability of certain connections being made [14]. The discriminator is aiming to get better at finding the fake images whereas the generator is aiming to get better at producing convincing images.
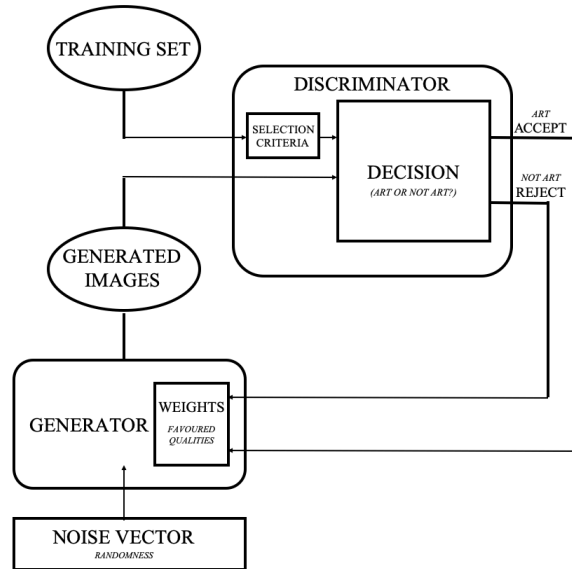

Figure 1 Generative Adversarial Network. Based on information from Goodfellow, I.J. et al. [13]

## 5 THE CREATIVE ADVERSARIAL NETWORK

The Creative Adversarial Network (CAN) works with the same basic premise as GANs. It consists of a generator and a discriminator [14]. The training set is also composed of images, however there is the addition of style labels (in Elgammal et al.'s original model, these were artistic style labels, such as 'abstract expressionism' [14]). This allows the discriminator to learn to distinguish not only what is or is not art, but also different categories of art.

The discriminator, as well as rejecting images which do not fit into its model of 'art', is also tasked with rejecting images produced by the generator that too closely fit into a specific style. The signal which is released by the discriminator to the generator is determined not only by whether the image is plausibly from the same set as the training (high scoring) but also whether the image can be unambiguously classified as one of the styles of art as introduced through the labelling of the training images (low scoring) or is more ambiguous (high scoring). This results in the generator tending towards stylistic ambiguity in art images it produces, whilst maintaining the qualities of artworks.

The generator in the CAN, as in GANs, is blind to the training set that is fed into the discriminator. It receives random input from the noise vector, which forms the initial basis for image production. The generator gradually adds weights (increasing likelihood of certain connections being made) to its image production algorithm based on the feedback of the discriminator.

Random noise continues to be inputted into the generator, which, combined with the "learning" from the discriminator's feedback, leads to the production of an image. The input of the noise vector ensures that any positively scored image is not merely repeated [14].

Figure 2 Creative Adversarial Network, based in part on
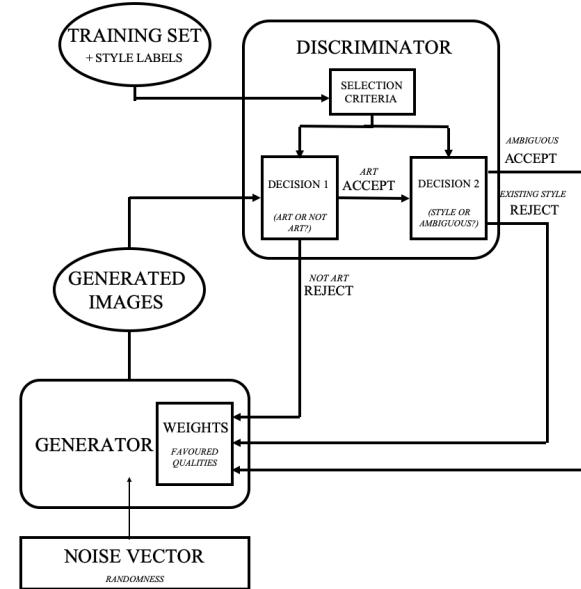
Elgammal et al. [14]

6 APPLYING THE DARWINIAN MODEL TO GANS AND CAN

GANs and the CAN can be shown to map onto the Darwinian model of creativity. Whilst they do not explicitly follow an evolutionary model (unlike genetic algorithms), they do inadvertently follow the model of creativity put forward by Simonton [1], suggesting that in the Darwinian sense at least, the two computational systems meet the criteria for creativity.

The key element of the Darwinian model of creativity is the non-teleological nature of the creation; the variation must occur without a view to what would be a successful mutation/recombination. This is the blind variation component of Simonton's model.

Blind variation is present in both CAN and GANs. The generator operates as the means of producing variation; it is not able to see the criteria of selection (what will be accepted rather than rejected), as this is derived by the discriminator from the training images. In this way, the generator is blind. The variation is ensured by the noise vector, which acts as a randomness generator, much like environmental stimuli in Simonton's model.

The success of the image is judged by the discriminator, as this controls the selection criteria. The discriminator compares the generated images to the selection criteria (which has been derived from the training set and style labels) and determine whether the image is successful or unsuccessful. This is the selection element of the model. The selection feedback from the discriminator can be understood as a generational change; the weighting added is much like the genes passed from generation

to generation. This equivalent to the retention of successful traits.

There is no huge difference between GANs and CAN in terms of Darwinian creativity, as the underlying process is the same and can be successfully mapped onto Darwinian creativity, there is not much concern for the distinction between the two. However, the CAN ensures that there will be ideational (in the case of art, stylistic) variation. As stated by Boden [7], ideational variation is vital for reaching near-human levels of creativity. Furthermore, in the case of other theories of creativity (which could complement the Darwinian model), the insurance of originality is of high importance [12]. This is only achievable in the CAN model, which ensures deviation from stylistic norms.

The diagrams below illustrate how the Darwinian model can be applied to GANs (fig. 3) and CAN (fig. 4).
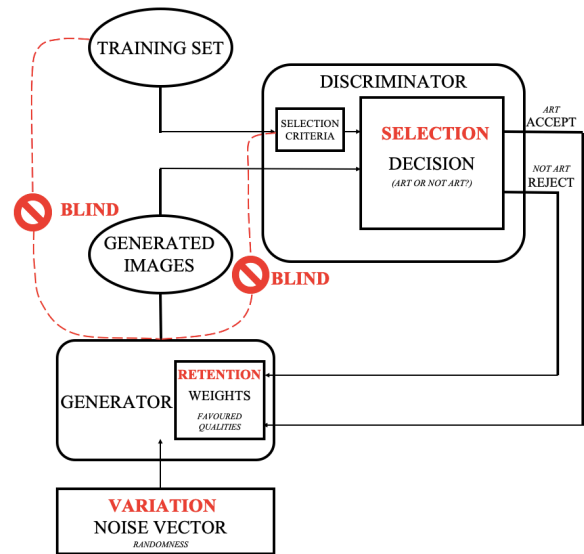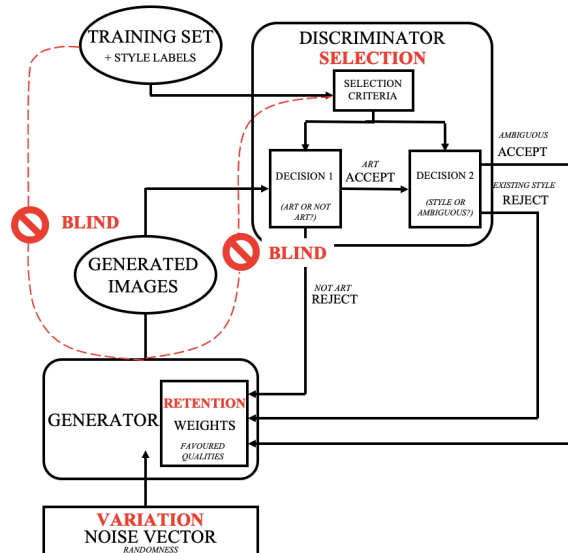

Figure 3 Darwinian model applied to GANs

Figure 4 Darwinian model applied to CAN

7 POTENTIAL OBJECTIONS

Objectors to the outlined argument may state that the application of the Darwinian model to these computer systems is merely an analogy, and therefore an argument that a computational system is creative because it maps onto the Darwinian model of creativity is merely making an analogy and does nothing to prove that the system is actually creative.

This is, however, exactly what is occurring in the Darwinian model of creativity as applied to any process; it is not specific to the application of machine creativity. It is a model which can be applied to other areas of thought analogously. If the model can be applied equally to the theory of creativity and machine "creativity", this suggests they are somewhat similar in functionality. In the case of developing computationally creative systems, a method of measuring some similarity to a human model of creativity is helpful in evaluating the system.

Some may also suggest counter-examples of non-creative processes, which could also be said to meet the terms of the model in the same ways as GANs and the CAN. However, as the Darwinian model is a broadly applicable model, this does not defeat the argument. Many processes may be found to fit with the Darwinian model. It is possible to question the utility of the Darwinian approach to creativity based on this, but this in itself is not a reason to protest the application to computational creativity.

Another objection to the proposed argument stems from an objection to the Darwinian model of creativity itself. This objections states that the Darwinian model is insufficient for creativity, and therefore meeting the criteria of this model is not enough to demonstrate creativity. This may be correct; however, I would suggest that the Darwinian model provides a *necessary* (though, perhaps not *sufficient*) condition to achieve creativity. Whilst this by no means proves outright machine creativity in the cases of GANs and CAN, their creativity cannot be ruled out based on not meeting the requirements of the model.

A final objection to the proposed argument may be from teleology. This objection would argue that both GANs and CAN fail to meet the requirements of Darwinian creativity as they are goal directed and therefore teleological. If sustained, this objection is potentially fatal to this argument as it proves the existence of false-analogy, removing the whole premise of Darwinian creativity being applicable to these computational models.

There are several potential responses to this objection. The first would be to deny that there is any intention or goal-directness in the systems as a whole, and therefore the objection is baseless. I will not pursue this course, as this would destroy in part the applicability of the whole Darwinian model to creativity, which is generally agreed to involve some level of intention [12]. A less problematic rebuttal would be to suggest that while the whole system is indeed somewhat goal directed, this does not mean that it does not fit the criteria of non-teleology of the Darwinian model. The components of the system are not goal-directed. Just as evolutionary mutation and recombination is not aimed at anything, neither are the generators in GANs and CAN; they are producing images based on the random noise, with the later addition of information of the retained qualities from feedback from the discriminator. As the generator meets the requirements of 'blindness' due to its lack of access to the training materials or selection criteria, it still cannot be said to know what it is aiming at.

8 THE DARWINIAN MODEL AS A TOOL FOR EVALUATION OF COMPUTATIONAL CREATIVITY

The application of the Darwinian model to GANs and the CAN shows that this model of creativity can function as a tool for assessing computationally creative systems

Unlike some models of creativity which have no clear measurability (such as Gaut's account [11] which requires agency, intentionality and understanding of values), the Darwinian model provides a clear way in which a system can be assessed to meet certain creative standards: it must include blind variation, selection, and retention of successful traits. With the added requirement of ideational variation, this model offers a measurable standard of creativity for computational systems.

While it may be the case that meeting the requirements of the Darwinian model of creativity is insufficient to be considered creative in the human sense, this model may provide a good initial method for assessing whether computationally creative systems meet some of the *necessary* conditions for creativity.

9 CONCLUSION

The model of creativity proposed by Simonton follows Darwin's evolutionary theory, which has since been used to model various psychosocial processes, including creativity. This model is comprised of blind variation, selection and retention, with the addition of ideational variation in the case of creativity, to ensure outputs are creative in a meaningful sense. Both GANs and the CAN can be successfully mapped onto Simonton's model of creativity. This suggests that these computational systems meet the standard of creativity laid out in the Darwinian model. Whilst this may not be sufficient to claim that GANs and CAN are creative, not meeting these criteria would have prevented them from being considered as such. This shows how the application of the Darwinian model can be used to assess computationally creative systems in a measurable way, unlike other popular theories of creativity. Whilst the Darwinian model may not be sufficient to prove creativity on a par with humans, it can provide an initial standard of assessment for computationally creative systems.

REFERENCES

[1] Simonton, D.K. Creativity as Blind Variation and Selective Retention: Is the Creative Process Darwinian? *Psychological Enquiry* 10:309–328 (1999)

[2] Campbell, D.T. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review,* 67(6): 380-400. (1960)

[3] Elgammal, A. et al. CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms. *arXiv* [Online]:1–22. (2017). Available at: http://arxiv.org/abs/1706.07068.

[4] Campbell, D.T. Evolutionary epistemology. In: Schlipp, P. A. ed. *The Philosophy of Karl Popper*. La Salle, IL: Open Court., pp. 413–463. (1974).

[5] Dawkins, R. *The Blind Watchmaker*. New York: Norton. (1986)

[6] Sternberg, R. J., & Davidson, J.E. *The Nature of Insight*. Cambridge, MA: MIT Press. (1995)

[7] Boden, M.A. *The Creative Mind: Myths & Mechanisms.* New York: Basic Books. (1991).

[8] Luke, S., Hamahashi, S. and Kitano, H. "Genetic" Programming. In: *GECCO'99 Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation* - Volume 2. pp. 1098–1105. (1999)

[9] Bäck, T. *Evolutionary Algorithms in Theory and Practice.* Oxford: Oxford University Press. (1996)

[10] Koza, J.R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: MIT Press. (1994)

[11] Boden, M.A. *The Creative Mind: Myths and Mechanisms.* 2nd (expan. London: Routledge. (2004)

[12] Gaut, B. The Philosophy of Creativity. *Philosophy Compass* [Online] 5:1034–1046. (2010). Available at: `http://doi.wiley.com/10.1111/j.1747-9991.2010.00351.x.`

[13] Goodfellow, I.J. et al. Generative Adversarial Networks. *arXiv.* (2014) [Online]:1–9. Available at: `http://arxiv.org/abs/1406.2661.`

[14] Elgammal, A. et al. CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms. *arXiv* (2017) [Online]:1–22. Available at: `http://arxiv.org/abs/1706.07068.`

Role-Based Perceptions of Computer Participants in Human-Computer Co-Creativity

Anna Kantosalo[1] and **Anna Jordanous**[2]

Abstract. The purpose of this ongoing research is to better understand the potential contributions that computers can play in situations where people interact with computers towards creative pursuits and goals. Past research has provided sets of definitions of different roles that a computer plays in human-computer creative collaboration. Thus far, we look into the advantages and limitations of having such roles. In particular, this paper contributes an analysis and categorisation of the coverage of existing role classifications for computational participants in co-creativity. This analysis is complemented by a comparative review of the use of roles to understand and structure creative collaboration between people only (i.e. without any computational participants involved). Our wider project investigates whether these defined sets of roles are a. adequate and b. helpful for understanding the perception of computational contributions in co-creativity, with a study planned to investigate the roles of current systems in practice. This project considers both co-creative computer systems that currently exist, and systems that could potentially exist in the future. Our goal is to reach a point where the perception of what is possible in human-computer co-creative collaboration is enabled and boosted (but not constrained) by a definitive set of roles.

1 INTRODUCTION

Computers are taking on more creative tasks in collaboration with humans. *Human-computer co-creativity* is a field which looks at the collaboration of computers with humans and other computers on a creative task. Such co-creative computer systems can contribute in an impressive array of different creative scenarios, from drawing to poetry (see e.g. [4] or [12]).

Alongside these practical achievements, we are developing a better theoretical understanding of how computers can contribute to co-creative scenarios. More specifically we ask, what roles can/do computers take, when collaborating with humans in interactive creative scenarios? This is an important route towards understanding the (actual) current contributions and limitations of co-creative systems, as well as in examining possible biases that humans may have in designing, implementing and evaluating different co-creative scenarios and the extent of collaboration in different scenarios [11].

'People do have a tendency to discount and even dislike computer creativity' [21, p.6]. Moffat and Kelly examined reactions to music composed by a computer system, and found that some musicians displayed (often subconscious) discrimination, reacting negatively to computer composed music. Similar biases arose (though not at a significant level) in Pasquier et al.'s experiments ten years later [26] and in the evaluation experiments in [10].

A reasonable concern is that the possible biases being introduced by human evaluators (and designers) are imposing additional limitations on co-creative systems; perhaps perceptions of what computers can (or cannot) do are influencing what we attempt with co-creative systems that interact with people. This is the issue we investigate in this project. Are our existing computational role classifications affected by general perceptions of what computational creativity capabilities, and if so, does this mean existing classifications are inadequate, over restrictive or limiting?

It is useful at this point to define some more key terms that we use during this paper. *Collaboration* is defined by [28] as: "a process in which two or more agents work together to achieve shared goals."

We define, in this paper, that a *role* is an assignation of specified responsibilities and behaviours that an agent plays in a collaboration. Roles can be emergent or pre-defined. In this paper we focus on pre-defined roles, as current computer roles are typically dictated in advance when building use-cases, and the analysis of emergent roles for the computer would require more data on long-term human-computer co-creative collaborations, which are scarce.[3]

Several role classifications exist for humans and computers collaborating in co-creative scenarios. We collate and analyse these below. Our concern is whether the coverage of existing role classification schemes is adequate and appropriate for describing and analysing the space of possibilities for computers and humans to collaborate co-creatively. An early example given by [31], describes how a computer can make strong contributions to co-creative scenarios through the ability to perform repetitive tasks accurately and rapidly to assist the creativity of the human participant(s). Their stated focus is on "the degree to which a PCG algorithm which generates valuable and novel content for the human designer to consider can contribute to human creativity." [31, p. 4]. This is indeed one useful type of role that computers can play in creative scenarios. A notable concern, however, is whether computer roles in such schemes overly focus on enabling creativity by the human participant(s), relegating the creative capacity of the computational agent(s) to that of a supporting role rather than as creative contributors in their own right.

What might we gain from more evenly balancing the creative responsibilities in human-computer co-creativity? Burleson [3] posits the potential benefits for a hybrid human-computer system, in that it could positively contribute to the creativity overall, and that of both the human and computer participants. A similar approach is also taken in more recent mixed-initiative creative interfaces (MICI) research (e.g. the MICI workshop at CHI 2017, where many papers "give the computer the status of creative agency and initiative thanks to AI"[6, p. 629]). As reported below, Kantosalo and Toivonen [13]

---

[1] Aalto University, Finland, email: anna.kantosalo@aalto.fi
[2] University of Kent, UK, email: a.k.jordanous@kent.ac.uk

[3] A new role for a computational system could emerge over time, for example, if it is used for a purpose it was not originally designed for.

have taken some steps towards modelling computational co-creative systems; our ongoing project moves beyond those steps to identify a set of roles for co-creativity with broader and more accurate, detailed coverage of the potential roles that a computer collaborator can take.

This paper's main contributions are: i) the analytical specification of the space of possibilities we get from existing human-computer co-creativity role classifications, and ii) the comparative analysis of these role classifications to the use of roles in human-human co-creativity scenarios.

This paper contributes to an ongoing wider project. Our project research question asks whether existing roles categorisations are sufficient for covering the entire space of possible roles that is possible with computational partners. We consider both co-creative computer systems that currently exist, and systems that could potentially exist in the future. Do the roles that we have in the current literature categorise the whole space of co-creative possibilities? Or, do they just cover what we have at the moment - or do they not even cover what we have at the moment?

This paper represents completion of the first stage of this work in progress; setting the theoretical foundations in place, so we can continue onto planned studies to evaluate existing roles in practice in current co-creativity systems. This will allow us to experimentally evaluate the coverage of existing roles, identify gaps in the coverage, and explicitly specify possibilities for new roles that are not yet considered in the literature.

Our goal is to reach a point where the perception of what is possible in human-computer co-creative collaboration is enabled and maximised (but not constrained) by a definitive set of roles useful for designing and analysing co-creative systems.

2 ROLE CLASSIFICATIONS IN HUMAN-COMPUTER CO-CREATIVITY

Possible roles for computers in the creative process have been presented in creativity support tool literature and more recently in computational creativity literature. We start by reviewing early role categorisations from creativity support tool literature and then consider more recent categorisations from literature focused on human-computer co-creativity.

We have listed the different roles and their origins in table 1, which also shows some overarching themes or categories between different role categorisations; we contribute these overarching themes as an output of our comparative analysis.

The rows are ordered by date (from oldest work to newest, moving from left to right), and the columns are ordered according to the autonomy and responsibility afforded to the creative agents (from least to most, as we move down the rows). It is interesting to note that as we move through time, from 2005 onwards, there is an increase in the maximum amount of creative responsibilities covered by roles in the classifications. In other words, over time, there is a general trend towards allowing computational participants increasingly more complex and more autonomous roles.

2.1 Supporting roles for computers in creative contexts

Many of the earlier role classifications for human-computer co-creativity focus on distinct supportive roles. This suggests the importance of support roles for creative collaboration.

Perhaps the most well cited description of possible roles for computers in the creative process comes from Lubart's 2005 [16] intro-

duction to the special issue of Computers in Human behaviour. He considers four distinct roles for the computer; The *Nanny*, which is a supportive role, encouraging the creativity of an individual human; The *Pen-Pal*, which is also a supportive role focused on facilitating communication between creative partners; The *Coach*, which considers increasing the human's creative capability through teaching, and finally; The *Colleague*, a computer which is able to aid a human by contributing new ideas. This classification is very much oriented to creativity support rather than active participation in the creative process. Only the last category, Colleague, allows for an active contribution from the computer.

Another classification of roles focused on the support for creativity rather than co-creativity comes from Nakakoji [24]. It includes three roles: The *Running shoe*, which focuses on supporting faster creation; the *Dumbbell*, which focuses on training human creative capacity; and the *Skis*, which describe the role of systems enabling completely new ways of creating, such as new instruments for musicians. These roles echo the two themes support and training in Lubart's classification and introduce a third, new category focused on enabling human creative behaviour.

The first role categorisation originating from computational creativity literature is by Maher [17]. It is still very much focused on the supportive roles a computer can have in the creative process, giving categories of *Support* and *Enhance*, which deal with providing tools and techniques for creativity or enhancing the creative capabilities of the human by encouraging creative cognition. Like Lubart, Maher, whose categorisation stems from the practical classification of interactive computationally creative systems, considers a more active role, the role of a *generator*, for the computer.

Some supportive roles are also echoed by Negrete-Yankelevich and Morales-Zaragoza [25], who note that in addition to being active participants in creative collaboration, computers can also support creativity by providing the general *environment* or *toolkits* for the creative humans.

From these suggested supportive roles we have derived three possible overarching categories for computer support: a general support role focused on traditional productivity aspects such as facilitating faster creation, a training role focused on teaching or training creative ability, or an enabling role focused on allowing for new[4] forms of creativity.

2.2 Participatory roles for computers in creative contexts

The role of providing new materials for the human to work on by generating new creative artefacts or parts of them is visible in many roles suggested for computers in the co-creative process, including not only Lubart's [16] and Maher's [17] classification, but also the classifications of Negrete-Yankelevich and Morales-Zaragoza [25], Kantosalo and Toivonen [14], and Hoffman [8]. But the generate-role is but one way in which a computer can be an active part in creative collaboration. In addition we have identified roles dealing with the evaluation of creative artefacts, problem finding, and ways to control the initiative in creative collaboration.

Some categorisations such as the categorisation by Kantosalo and Toivonen [14] give distinct roles for concept generation, evaluation and definition, but for example Negrete-Yankelevich and Morales-Zaragoza [25] have given a role definition, in which the roles are

---

[4] By 'new forms of creativity' we include both P-creativity (new for the creator) and H-creativity (completely new) [2].

CATEGORISATIONS [this paper]		Lubart, 2005 [16]	Nakakoji, 2006 [24]	Maher, 2012 [17]	Negrete-Yankelevich & Morales-Zaragoza, 2014 [25]	Kantosalo & Toivonen, 2016 [14]	Guckelsberger et al., 2016 [7]	Hoffman, 2016 [8]
Creativity Support Tools	SUPPORT	Nanny	Running shoe	Support	Environment			
		Pen - Pal			Toolkit			
	TRAIN	Coach	Dumbbell					
	ENABLE		Skis	Enhance				
Co-Creative Colleagues	GENERATE	Colleague		Generator	Generator	Generator		Divergent agent
					Apprentice			
	EVALUATE					Evaluator		Convergent agent
					Master			
	FIND PROB-LEMS					Concept definer		Divergent agent
	CONTROL INITIATIVE					Pleasing Agent	Supportive Agent	
						Provoking Agent	Antagonistic Agent	

Table 1. Roles for computers in the creative process, including overarching categories that our analysis has produced.

additive, describing increasing capabilities of the computational collaborator: A generator can only generate, whereas an apprentice is also able to evaluate its output, and a master can also decide what to create to some extent. Likewise Hoffman's [8] categories of divergent and convergent agents stemming from the theory of divergent and convergent thinking are roles that do not deal with just a specific task or ability.

Finally, in addition to the ability to participate in specific tasks required by the roles of generate, evaluate and find problems, there are roles that better fit a behavioural strategy selected by the computer in human-computer co-creation. Both Kantosalo and Toivonen [14] and Guckelsberger et al. [7] have suggested that the computer could try to actively satisfy their human collaborators needs by assuming a pleasing or supportive role, or more actively challenge the human or focus on its own goals by assuming a provoking or antagonistic role.

In addition to the roles presented here there are many domain specific roles that have not been discussed in this paper as they are not generally applicable.

3 ROLE CLASSIFICATIONS IN HUMAN-HUMAN CO-CREATIVITY

Having defined the space of possibilities covered in existing role classifications for human-computer co-creativity, we turn our attention to role classifications in human-human co-creativity. As part of this paper's contribution, we will now consider the use of role classifications for scenarios where people collaborate with other people in creative tasks.

We are familiar with the concept of assigning roles to people working in teams or other collaborative scenarios. Arguably the most well-known examples of sets of roles that humans take (consciously or unconsciously) in collaborative scenarios are the Belbin team roles for collaboration [1]:

- *Plant* - idea generators, innovators
- *Resource Investigator* - networkers, creating opportunities
- *Co-ordinator* - organising people and focusing on the bigger picture
- *Shaper* - task-focused achievers
- *Monitor Evaluator* - critical observers
- *Teamworker* - those who maintain or fix group relations
- *Implementer* - those who put plans into action
- *Completer Finisher* - getting things done as well as possible
- *Specialist* - experts in an area of knowledge

While there is scope for creative results to be achieved using teams organised around Belbin's roles, arguably the only role with direct creative responsibilities is the *Plant* role. Our interests here lie more in teams comprised of creative individuals (human and computational).

Edward de Bono proposed a theory around different thinking styles which could be applied to our goal of roles in creative collaboration. The six distinct thinking styles identified by de Bono are

represented as different coloured *Thinking Hats* [5]:

- Blue hat thinking - process
- Red hat thinking - feelings
- White hat thinking - facts
- Yellow hat thinking - benefits
- Black hat thinking - cautions
- Green hat thinking - creativity

These 'hats' could be relevant as different roles for co-creative participants, as de Bono considered their application to complex problem solving, which is often seen as an arena for creativity; de Bono argued that people could wear different hats to approach about a problem in different ways (though an individual person can swap between hats as needed). There is less focus on how different hats interact in collaboration, which is something we wish to uncover in our roles. Also, we have a similar issue to the Belbin team roles, in that only the Green hat is specifically associated with being creative.

Bringing us closer to the remit of human-computer co-creativity, Mamykina, Candy and Edmonds [18] talk about support for human-human co-creativity in the context of an artist working with a technologist. They found several different models for co-creativity: *assistant model*, *full partnership* and *partnership with artist control*. The models differ in terms of who does what, so the choice of model depends on what tasks an artist and technologist assume during collaboration (with the possible tasks being: *creative concept*, *construction* and *evaluation*). In the assistant model, the artist is responsible for the creative conceptualisation (initial idea) and evaluation, with the technologist responsible for the construction. In the full partnership model, both do all tasks, and in the model of partnership with artist control, both do concept creation and construction, but only the artist evaluates. Note that the artist is always afforded responsibilities for creative conceptualisation and evaluation, but not always for construction. Similarly, the technologist is always afforded responsibilities for construction, but is restricted in most models from evaluative responsibilities and in one model from creative conceptualisation. The roles, tasks and resulting models are interesting in terms of understanding different scenarios that might be encountered in human-computer co-creativity, but the work in [18] does not necessarily bring us any closer to identifying different types of roles beyond 'artist' and 'technologist'. What types of participant comprise a creative team?

Modelling the composition of a group in team-based creativity is discussed by Reiter-Palmon, Wigert and de Veerde [27], who review various comparisons between heterogenous teams (with participants of various types) and homogenous teams (with participants similar in type). Observations arise on on the usefulness of 'functional diversity' (i.e. participants offering different functionality to the team), however no detail is given on what actual functions might be useful to include.

West [30] proposes a theory of creativity in work groups that includes investigation of team diversity along three other factors: task, team integration and external demands. In his work it is apparent how roles are very important for social systems, but they are seen more as a fact arising from occupational constraints and responsibilities rather than roles established for the creative activity itself.

Mumford, Whetzel and Reiter-Palmon [23] make similar observations on roles arising from occupational requirements, although they are looking at organisational creativity (specifically, the emergence of creativity in organisations over time), as opposed to creative collaboration per se. Mumford et al. include a thorough definition for roles in organisations, and describe how roles relate to creative problem solving: "Role requirements and role characteristics are not defined in an arbitrary fashion. Instead, role requirements and role characteristics emerge, in part, as a function of the issues confronting the organization." In other words, roles emerge relevant to the scenario rather than being defined independently of the scenario being tackled. Mumford et al. go on to say that "even when the requirements of a role call for creative thought, the nature and success of peoples' creative problem solving efforts may be conditioned by other characteristics of their roles." Hence participants in creative collaboration are not purely defined by their creative contributions, but by broader behaviours and motivations.

It has been argued that assigning predefined roles to participants actually *stifles* creativity [19]. This is a very important point for our purposes. Wang, Xhang and Martocchio [29] investigated the alternative of role *ambiguity* and its effects on creativity, hypothesising that too little or too much ambiguity in the definition of roles both limit creativity (but tolerance to such ambiguity is useful). They presented experimental evidence supporting this hypothesis, suggesting that moderately defined roles are important for human creativity in organisations. For our purposes, this suggests that defining roles for the human and the computer is important for the successful adoption or analysis of the co-creative system, but that a strict role classification may actually constrain creativity. Similar observations have been found in the study of how constraints affect computational creativity [20]; too many constraints or too few constraints had negative repercussions for the level of creativity exhibited by the system. Perhaps, in future work when we are evaluating existing role classification systems, useful parallels can be drawn between a. defined roles for creative participants and b. defined constraints that creative participants must work under?

In summary: we do see examples of specific roles being assigned to people participating together in team work, such as Belbin's team roles [1] or de Bono's thinking hats [5]; however these often restrict the creativity to being part of only one participating role (or hat), rather than creative contributions coming from multiple participating roles. Mamykina et al. [18] hint at a similar restriction of creative responsibilities in their thoughts on artist-technologist collaborations. Team heterogeneity can be achieved by introducing different roles, with positive effects on creativity [27, 30]. Mumford et al. [23] advocate allowing a more fluid set of roles to emerge over time. Imposing overspecified predefined roles may, however, overly constrain the creative potential of the team [19, 29].

Having reviewed the use and applicability of roles in creative collaborations with people, we can now use this knowledge to recontextualise and reconsider the existing role classification systems for human-computer co-creativity. We do this by examining parallels that can be drawn between human-computer and human-human co-creativity roles, in the next section. Then we discuss more generally what information we gain for our understanding of computational participants' roles in co-creativity.

4 DISCUSSION

We now have a clearer specification of the existing role classification systems for human-computer co-creativity, and knowledge about the use of roles in human-human co-creative scenarios. What information do we gain from comparing this knowledge, about the perceived and actual range of possibilities for computational contributions in human-computer co-creativity?

4.1 Parallels between human-computer and human-human co-creativity roles

In both human-computer and human-human co-creativity role classifications we see classifications that focus on more task-based roles, e.g. [16, 25, 17] and [18] respectively; we also see classifications oriented towards including more complex behaviours e.g. [8, 14, 7] and [5, 1] respectively.

There are similarities between Hoffman's [8] divergent and convergent thinker roles and de Bono's [5] thinking hats, which become strategies to drive divergent and convergent thinking. Both role classifications require the agent to adopt a specific stance to the problem at hand and deploy different ways of thinking.

There are also similarities between Mamykina et al.'s models [18] and Negrete-Yankelevich's and Morales-Zaragosa's [25] roles. Mamykina et al.'s assistant model corresponds with the computer participant taking a Generator role in Negrete-Yankelevich's and Morales-Zaragoza's terms (with the computational participant as technologist and human participant as artist); full partnership corresponds to the computer taking the master role (where human and computational participants could either be assigned as technologists or artists). The only model in [18] which does not fit quite so neatly with [25]'s roles is the partnership with artist control model, which falls in between the Apprentice and the Master roles for the computer.

Overall, there are parallels between the distinct models proposed by Mamykina et al. and some of the overarching divisions that we have identified from comparing various human-computer collaborative roles: specifically, the *Generate*, *Evaluate* and *Find problems* divisions.

It is naturally easier to equate the artist in [18] to the human participant, and the technologist to the computational participant. But if the reverse assignment was considered (the artist is the computational participant, the technologist is the human participant), then the Mamykina et al. models help open up thinking about the possibilities of more diverse roles for computational participants in co-creativity. Currently, for example, there is no role in the classifications reported above in which a computer participant outsources generation but retains evaluative control and problem definition on its own. This is in part captured by the model by Kantosalo and Toivonen [14], which suggests that in task-divided co-creativity creative responsibilities can be divided in this way, however for this the specific roles in the model have to be considered as additive roles.

Although the Belbin [1] team roles are more for general team collaboration instead of specifically co-creative collaboration, it is interesting to attempt to draw parallels between the Belbin roles and the categorisations that have emerged in our comparative analysis of human-computer co-creativity roles. Some possibilities have been mapped below:

- Plant → FIND PROBLEMS
- Resource investigator → ENABLE
- Co-ordinator → SUPPORT; CONTROL INITIATIVE
- Shaper → GENERATE
- Monitor evaluator → TRAIN; EVALUATE
- Teamworker → SUPPORT; TRAIN; ENABLE
- Implementer → GENERATE; CONTROL INITIATIVE
- Completer Finisher → GENERATE; EVALUATE; CONTROL INITIATIVE
- Specialist → ALL [depending on how they use their knowledge]

As shown above the Belbin [1] roles can be mapped into several

categories in our analysis. This suggests that the Belbin roles are more nuanced than the typical roles assigned in human-computer co-creativity to computational collaborators. Computers can of course act in multiple roles simultaneously, but it is worth considering if our categorisations should name some of these unique combinations, producing something more similar to Belbin. This type of thinking is already somewhat present in the additive roles by Negrete-Yankelevich and Morales-Zaragoza [25].

Current human-computer co-creativity roles have little variety in terms of the 'control initative' roles, such as leadership[5], which seem to emerge in Mamykina et al.'s [18] roles, as well as in Belbin's [1] team roles. This moves us towards considering differences between human-human and human-computer co-creativity roles.

4.2 Differences between human-computer and human-human co-creativity roles

No standard role classification system has emerged from our review of human creative collaboration. A reoccurring observation is that over-prescriptive or overly general role characterisations perhaps stifle rather than support creativity. This is a very important point for our purposes. Perhaps our attempts to categorise role classifications are misguided and unnecessary?

We note, however, that having no accepted standard role classification system at present does not result in the situation where computers are given free rein in co-creative scenarios. As discussed above, a computer's role in co-creativity is often limited by perceptions.

By failing to identify roles that we currently have and roles that we could have in the future, we are standing still. We would fail to have this analytical tool that we could use to better understand relationships between human and computational participants.

Hence we continue to pursue the use of roles as a model; this becomes a tool to analyse and better understand human-computer co-creative applications and potential. It also helps us to map the potential contributions that computers can make in co-creative scenarios, and identify areas that are under-explored, unlocking the potential of future research.

The ideal scenario would be to arrive at a point where perception-based limitations are no longer an issue, and hence the roles that we investigate are no longer needed. Until we reach that point, though, we argue that a more comprehensive role classification helps justify and positively emphasise the capabilities and potential contributions of co-creative computational systems.

5 FUTURE WORK AND EVALUATION

The existing role classifications that we have for human-computer co-creativity are theoretical, in the sense that they have not been experimentally verified. They are essentially hypothetical labels, assigned rather than analysed. In future work we will evaluate the current role classifications as they are manifested in practice, which will help us to identify gaps and coverage of the existing classifications.

We plan to gather evidence through a study with computational creativity / human-computer interaction researchers. The participants will be given several scenarios describing use-cases of co-creative systems. The participants will also be given a list of possible roles for participants in the co-creative scenarios, accompanied by brief descriptions of each role. These roles will be drawn from the above reviewed literature. Each participant will be given each scenario one

---

[5] The role of leaders is according to [22] typically seen as passive, guiding, but they suggest the leader is in fact an active collaborator.

by one, in randomized order, and they will be asked to rate how strongly each of the possible roles describe each of the collaborators indicated in the current scenario. In addition, the participant may suggest possible new roles in an open field.

We note two points that are particularly important to control for.

1. Firstly, the creativity of the scenario itself is an important question to ensure the scenario falls within the co-creative space (i.e. the creative domain in question may play a role in people's perceptions of what is creative, for example someone may think painting a picture is more creative than writing computer software – or vice. versa). Hence participants will be divided into two groups; the included scenarios will either be presented in domain-independent language (group 1), or will cover multiple creative domains representative of current computational creativity research, as guided by the mapping of research in [15] (e.g. musical creativity, linguistic creativity, etc) (group 2).

2. Computer creativity can be a controversial question to consider. This issue must be acknowledged in this work, as in any computational creativity research, but detailed investigation is beyond the scope of this work. There are numerous works looking into this question e.g. [9]; to control for this concern, we specifically target participants who are familiar with computational creativity research, and we will write the scenarios in neutral language such that it is not stated which is the computational partner and which is the human partner. Participants in the study will be made aware, for full disclosure, that they are evaluating scenarios which describe examples of human-computer co-creativity.

If we gather sufficient evidence that existing roles offer inadequate coverage of the space of possibilities for computers in co-creative scenarios, our research will then develop new theoretical models of computational roles, informed by our participants' comments. This model is intended to influence the ways people working in this field operate in future, by broadening perception of what computers can be capable of in co-creativity. We also hope to validate the theory in practise by observing human-human and human-computer pairs working on the same tasks.

6 CONCLUSIONS

In this paper, we outlined and categorised the space of co-creative possibilities covered by the current sets of roles in the literature. We looked for inspiration from reviewing how role classification schemes are used to understand and analyse creative collaboration between people, and we reflected on implications from this review.

Role classifications in co-creativity have appeared over recent years and have increasingly afforded computational co-creativity participants with more creative agency and more varied responsibilities. We seem to be moving towards a less limited set of views on what computational participants can do, though we are not yet at a point where computational participants are given the same agency as human participants.

In introducing extra roles to support the perception of computers in a wider range of creative behaviours, our end goal actually is to reach a point that human-human collaborative creativity is at, i.e. where predefined generally applicable roles are no longer necessary at all. This seems somewhat contradictory; however we argue that roles have emerged as an important analytical tool to study the potential of computational partners in collaborative creativity. Roles can also act as prompts to identify and highlight underexplored possibilities: becoming stepping stones towards more creativity potential.

REFERENCES

[1] Belbin, *Team Roles at Work*, Routledge, may 2012.
[2] Margaret A. Boden, 'Creativity and artificial intelligence', *Artificial Intelligence*, **103**(1), 347 – 356, (1998). Artificial Intelligence 40 years later.
[3] Winslow Burleson, 'Developing creativity, motivation, and self-actualization with learning systems', *International Journal of Human Computer Studies*, (2005).
[4] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko, 'Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent', in *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 196–207, (2016).
[5] Edward De Bono, *Six Thinking Hats*, Little Brown and Company, 1985.
[6] Sebastian Deterding, Jonathan Hook, Rebecca Fiebrink, Marco Gillies, Jeremy Gow, Memo Akten, Gillian Smith, Antonios Liapis, and Kate Compton, 'Mixed-initiative creative interfaces', in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, p. 628–635, New York, NY, USA, (2017). Association for Computing Machinery.
[7] Christian Guckelsberger, Christophe Salge, Rob Saunders, and Simon Colton, 'Supportive and antagonistic behaviour in distributed computational creativity via coupled empowerment maximisation', in *Proceedings of the seventh international conference on computational creativity*, (2016).
[8] Oliver Hoffmann, 'On modeling human-computer co-creativity', in *Knowledge, Information and Creativity Support Systems*, eds., Susumu Kunifuji, George Angelos Papadopoulos, Andrzej M.J. Skulimowski, and Janusz Kacprzyk , pp. 37–48, Cham, (2016). Springer International Publishing.
[9] Anna Jordanous, 'A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative', *Cognitive Computation*, **4**(3), 246–279, (2012).
[10] Anna Jordanous, *Evaluating computational creativity: a standardised procedure for evaluating creative systems and its application*, Ph.D. dissertation, University of Sussex, 2012.
[11] Anna Jordanous, 'Co-creativity and Perceptions of Computational Agents in Co-creativity', in *Eighth International Conference on Computational Creativity (ICCC'17)*, eds., Ashok Goel, Anna Jordanous, and Alison Pease, pp. 159–166, Atlanta, GA, (2017). ACC.
[12] Anna Kantosalo, Jukka M Toivanen, and Hannu Toivonen, 'Interaction Evaluation for Human-Computer Co-creativity: A} Case Study', in *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)*, eds., Hannu Toivonen, Simon Colton, Michael Cook, and Dan Ventura, pp. 276–283, Park City, Utah, (jun 2015). Brigham Young University.
[13] Anna Kantosalo and Hannu Toivonen, 'Modes for Creative Human-Computer Collaboration: Alternating and Task-Divided Co-Creativity', *Proceedings of the Seventh International Conference on Computational Creativity*, (2016).
[14] Anna Kantosalo and Hannu Toivonen, 'Modes for creative human-computer collaboration: Alternating and task-divided co-creativity', in *Proceedings of the seventh international conference on computational creativity*, pp. 77–84, (2016).
[15] Róisín Loughran and Michael O'Neill, 'Application Domains Considered in Computational Creativity', in *Eighth International Conference on Computational Creativity (ICCC'17)*, eds., Ashok Goel, Anna Jordanous, and Alison Pease, pp. 197–204, Atlanta, GA, (2017). ACC.
[16] Todd Lubart, 'How can computers be partners in the creative process: Classification and commentary on the Special Issue', *International Journal of Human Computer Studies*, (2005).
[17] Mary Lou Maher, 'Computational and Collective Creativity: Who's Being Creative?', in *Proceedings of the 3rd International Conference on Computer Creativity*, eds., Mary L Maher, Kristian Hammond, Alison Pease, Rafael Pérez, Dan Ventura, and Geraint Wiggins, pp. 67–71, Dublin, Ireland, (may 2012).
[18] Lena Mamykina, Linda Candy, and Ernest Edmonds, 'Collaborative creativity', *Commun. ACM*, **45**(10), 96–99, (October 2002).
[19] Mitch McCrimmon, 'Teams without roles: Empowering teams for greater creativity', *Journal of Management Development*, **14**(6), 35–41, (1995).
[20] Lewis McKeown and Anna Jordanous, 'An evaluation of the impact of constraints on the perceived creativity of narrative generating soft-

ware', in *9th International Conference on Computational Creativity*, eds., François Pachet, Anna Jordanous, and Carlos León, pp. 16–23. Association of Computational Creativity, (2018).

[21] David C Moffat and Martin Kelly, 'An investigation into people's bias against computational creativity in music composition', in *Proceedings of the 3rd International Joint Workshop on Computational Creativity (ECAI06 Workshop)*, Riva del Garda, Italy, (2006).

[22] Michael D Mumford, Shane Connelly, and Blaine Gaddis, 'How creative leaders think: Experimental findings and cases', *The Leadership Quarterly*, **14**(4-5), 411–432, (2003).

[23] Michael D Mumford, Deborah L Whetzel, and Roni Reiter-Palmon, 'Thinking creatively at work: Organization influences on creative problem solving', *The Journal of Creative Behavior*, **31**(1), 7–17, (1997).

[24] Kumiyo Nakakoji, 'Meanings of Tools, Support, and Uses for Creative Design Processes', in *Proceedings of International Design Research Symposium'06*, (2006).

[25] Santiago Negrete-Yankelevich and Nora Morales-Zaragoza, 'The apprentice framework: planning, assessing creativity', in *Proceedings of the Fifth International Conference on Computational Creativity*, eds., Simon Colton, Dan Ventura, Nada Lavrač, and Michael Cook, pp. 280–283, Ljubljana, Slovenia, (jun 2014).

[26] Philippe Pasquier, Adam Burnett, Nicolas Gonzalez Thomas, James B Maxwell, Arne Eigenfeldt, and Tom Loughin, 'Investigating listener bias against musical metacreativity', in *Proceedings of the Seventh International Conference on Computational Creativity*, pp. 42–51, (2016).

[27] Roni Reiter-Palmon, Ben Wigert, and Triparna de Vreede, 'Team Creativity and Innovation', in *Handbook of Organizational Creativity*, 295–326, Elsevier, (2012).

[28] Loren G. Terveen, 'Overview of human-computer collaboration', *Knowledge-Based Systems*, **8**(2), 67 – 81, (1995). Human-computer collaboration.

[29] Shuhong Wang, Xiaomeng Zhang, and Joseph Martocchio, 'Thinking outside of the box when the box is missing: Role ambiguity and its linkage to creativity', *Creativity Research Journal*, **23**(3), 211–221, (2011).

[30] Michael A West, 'Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups', *Applied psychology*, **51**(3), 355–387, (2002).

[31] Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos, 'Mixed-Initiative Co-Creativity', *Proceedings on the Foundation of Digital Games 2014*, (2014).

Will the real artist stand up? Computational creativity as mirror to the human soul

Joel Parthemore[1]

Abstract. This paper argues that a too-expansive view on creativity is unhelpful at best and deeply misleading at worst. As with "representation", the word "creativity" comes value-laden in ways that researchers cannot lightly get away from, if they can escape at all; simply claiming that one is using the word in a technical sense is not a solution. Neither should one take an overly narrow view that takes advantage of *a priori* arguments to deny creativity to classes of agents or putative agents solely by their membership in those classes. The paper proceeds by offering a definition of creativity meant to prejudice neither human being nor artefact; then setting out the conditions for a putative creative agent to be a creative agent, concluding that no existing artefactual agents appear to fall into this category; finally, addressing the question of why computers, computer programs, robots, and related artefacts have nevertheless had a profound – indeed, transformational – effect on human creativity, taking creativity to places that neither human beings nor artefacts could have gone on their own. It ends with a discussion of the person I see as one of the key early voices on computational creativity.

1 What is creativity?

> [Consider] an atomic pile of less than critical size: an injected idea is to correspond to a neutron entering the pile from without. Each such neuron will cause a certain disturbance which eventually dies away. If, however, the size of the pile is sufficiently increased, the disturbance caused by such an incoming neutron will very likely go on and on increasing until the whole pile is destroyed. Is there a corresponding phenomenon for minds, and is there one for machines? There does seem to be one for the human mind. The majority of them seem to be 'sub-critical'.... A smallish proportion are super-critical. An idea presented to such a mind may give rise to a whole 'theory' consisting of secondary, tertiary and more remote ideas... [30, p. 454].

> **Computational creativity:** The philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviors that unbiased observers would deem to be creative [9, p. 21].

There is a general consensus that creativity is one of the key aspects of human or human-like intelligence [32, p. 450]. That is about as far as the agreement goes, however. In particular, one finds a curious divide in the research community between those would insist that artefactual agents could never be creative, even in principle; and those who insist that they not only can be but already are.

In the former camp I would place Lund University's Jordan Zlatev, for whom living organisms are distinguished by their possession of intrinsic value, while certain organisms (notably enculturated human beings) distinguish themselves from others by their capacity to create meaning for themselves [34]. Artefacts – at least as they exist today – are not alive; therefore they cannot be creative; but Zlatev seems hostile to the possibility that they ever *could* be alive/creative, either. In the latter camp one finds no less a heavyweight in the computational creativity community than Maggie Boden, by whose definition of creativity[2] any number of virtual and physical artefacts[3] – not least Harold Cohen's AARON [7] – qualify as creative agents. The disagreement is not just or even particularly one of definition, however, but goes to the heart of what it means to be human.

Nevertheless, a working definition is required. I propose the following, which I will attempt to justify over the course of the paper:

> **Creativity.** The at least partly – yet never fully! – intentional act of an intentional agent or agents recombining elements of past or present experience in more or less strikingly novel ways to yield insights – from the subtle to the life- or world-altering – or[4] more immediate practical benefit.

A number of aspects of this definition require clarification:

1. ***at least partly:*** Unlike serendipity from which it usefully may be distinguished, creativity does not come about by accident. It cannot be the product of random chance (even though chance may play an important role); rather, it requires someone (call her the *creative agent*) trying to be creative. What this implies, too, is that it's not just the product that counts – a common metric in computational creativity – but the process by which it is produced [8].

---

[2] "...The ability to come up with ideas or artefacts that are new, surprising, and valuable" [1]; *cf.* Geraint Wiggins [32, p. 451]: "the performance of tasks which, if performed by a human, would be deemed creative."

[3] ...By which I mean anything that has been artificially constructed as the product of human activity rather than coming together as the result of naturally occurring processes.

[4] To be clear, the "or" is intended as "inclusive or".

[1] Guest researcher, University of Skövde, Sweden. Email: **joel.parthemore@his.se**.

2. **yet never fully! – intentional:** Neither can creativity be forced. A creative agent cannot simply decide to be creative, hence the well-known phenomenon of "writer's block" recognizable across all areas of creative activity.

3. **act:** The act need not be limited to the "outwardly" observable: an artwork, an invention. It might be the seeking out and discovery of an idea – one that could, but need not, be shared with others. Note that acts, by their nature, are never truly instantaneous ("in the moment"); they play out over time, and where one act stops and another begins may be difficult or impossible to say.

4. **of an intentional agent or agents:** The definition sets high requirements on agency, with the explicit aim of avoiding one rule for human beings, another for artefacts. The creative agent must be the sort of agent who can operate with conscious intent.[5] An agent lacking this capacity can certainly produce something that appears to be creative but the creativity, if any, will not be located in that agent. The requirements on creative agency will be discussed at length in Section Two. For now, note that the creative act need not the the product of a single creative agent's actions.

5. **recombining elements of past or present experience**[6]**:** Like imagination, creativity does not come out of nowhere, despite occasional appearances to the contrary; in some important sense, it is never truly new (or, as Boden writes, "there's new – and there's new" [1]). This definition emphasizes what Boden sees as only one type of creativity ("unfamiliar combinations of familiar ideas" [1]) at the apparent expense of the other two: exploratory and transformational [2]. What Boden sees as distinct types of creativity[7] this definition views as different aspects of a single phenomenon, aspects that may be more or less prominent depending on the circumstances but never fully absent. The exploratory aspect ("someone trying to be creative") has already been mentioned. The recombinatory aspect, however, is in many ways the cornerstone.

6. **in more or less strikingly novel ways to yield insights – from the subtle to the life- or world-altering:** Creativity comes in degrees, both in terms of how it is viewed at the time and how it is subsequently viewed, looking back. After all, a creative act that seemed relatively unimportant at the time may later be seen to have extraordinary consequences. Consider that the true significance of Johannes Kepler's creative improvements on Tycho Brahe's models of the solar system were not immediately apparent even to him – yet they came to play a critical role in the establishment of Nicholaus Copernicus' model as the accepted one in science. Here is the transformational aspect of creativity at its clearest.

7. **– or more immediate practical benefit:** The definition returns, at its close, to the mundane. A newly written haiku may serve simply to lift its author from a moment of depression. A new kind of hook may serve no greater purpose than allowing one particular picture to hang on one particular wall where it could not hang before. Still, even the smallest acts of creativity have a transformative role to play.

Note what this definition does *not* address: Boden's distinction between "psychological" (P) and "historical" (H) creativity, where P-creativity is new for the agent (or, though Boden does not seem to address this possibility explicitly, agents) doing the creative act, while H-creativity is, so far as anyone can tell, new for anyone, anywhere, at any time up till the present [2, p. 30]. Partly this is because the "so far as anyone can tell" qualification is playing a really critical role, especially if one opens up the possibility of other species on other worlds.[8] Partly it is because Boden seems open to the creatively new, particularly of the transformational variety, that does *not* involve recombining elements of past or present experience: creativity that really does, as it were, come from "nowhere".

It is a frequent observation that the "new" is often not so new as it appears or as people are quick to judge it. It is not quite as the writer in Jewish and Christian traditions puts it: "What has been is what will be, and what has been done is what will be done; and there is nothing new under the sun" [18, p. 805]; but even acts of creation that were not possible, in human terms at least, more than a few years or decades ago nevertheless stood on the shoulders of *other* acts of creation, the former blending into the latter in a broader arc of creative endeavor. Whereas Boden sees H-creativity as a special instance of P-creativity, I see it simply as a different perspective on the same phenomenon: one that, ironically, fails to put the creative act into its proper historical perspective and acknowledge its debt to the past. All creativity just is (individual or collective) P-creativity, in the end.

Note, too, that this definition is agnostic about the possibility (and value) of distinguishing different types of creativity, other than to resist Boden's distinctions (and those who would build on them). Neither is its purpose to set out the different levels of creativity: what counts as less or more creative. Such matters lie beyond the scope of this paper.

The next section attempts to apply the definition, reflecting on its consequences toward determining who counts as a creative agent – concluding that (in general) humans and members of some other species do, while existing artefacts do not. Section Three then addresses the question of how and why computational creativity remains a critically important endeavour: where it came from and where, perhaps, it is going.

---

[5] Alan Turing – inspiration for the final section of this paper – might be seen to be arguing against this with his rejection of what he calls the "argument from consciousness" [30, pp. 445-447] against the possibility of thinking (and, by extension, creative) discrete-state machines. Closer examination however shows that his target is solipsism and its tempting argument that "no one is conscious except me." For Turing, what later would come to be called *philosophical zombies* are, indeed, conceivable, but uninteresting.

[6] For an extensive discussion of the critical role of experience in creativity – albeit without a clear definition of what "experience" means – see [15].

[7] . . . Even as other researchers (e.g., Wiggins [32, p. 454]) have noted the vagueness of the boundary between them.

[8] Of course no one – on any world, presumably – created the Mona Lisa before da Vinci did; but da Vinci created it by drawing on existing painting techniques, along with his experience of human anatomy, along with – most likely – a model posing for him.

2 Who qualifies as a creative agent?

> ...Creativity is not a special talent possessed only by a fortunate elite. On the contrary, it is an unavoidable aspect of normal intelligence... [2, p. 41].[9]

> We live to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be *necessarily* superior, for then there is no danger of him losing his commanding position [30, p. 444].

2.1 Consciousness required

I have said that a creative agent must – by the definition offered here – be one capable of acting with conscious intent: that is, the creative agent must, at some level, be aware of what she is doing.[10] That probably rules out a range of other species (though hardly all, unless one is inclined again to *a priori* arguments and willing – as e.g. Zoltan Torey is [29, p. 90] – to discount the evidence of various species passing versions of the so-called *mirror test* [12] and evidence from corvids of creative tool construction and use: see e.g. [17]). It probably also rules out newborn human infants. Does it rule out artefactual agents like AARON? It would seem it does.

In [24], Blay Whitby and I set out to address and refute claims by a number of researchers – notably Uma Ramamurthy and Stan Franklin [27] – that their creations had achieved at least some degree of consciousness. The problem was *not* that they were artefacts; an earlier paper of ours [23] concluded that what we called the *artefactual question* was a red herring, confusing matters by suggesting that different rules might apply to "natural" and artefactual agents. *Pace* John Searle [28], it does not matter how the agent is built "inside" or even, ultimately, whether the agent's behaviour may be seen to be guided by an explicit list of rules – only whether the agent's behaviour is measurably different, in any relevant way, from that of any other agent to whom one would attribute conscious intelligence: i.e., a mind.[11] I have argued [22] that the hallmark of consciousness – as with conceptual agency (the two may be seen as two sides of a coin) – is a flexibility of behaviour in "reasoned" response to one's environment, allowing one to take one's past experiences into account in deciding how to respond to present circumstances[12]; while the

hallmark of intelligence is arguably (but non-controversially, I think) its independence from any particular domain of application. The problem with existing artefacts is that they lack that flexibility of behaviour and domain-independent reasoning. The problem with existing claims to conscious artefacts, we found, is that either the authors are reducing consciousness to some simpler value x, or they are coy about whether they are to be taken at face value or metaphorically.

An inspiration for both papers [24, 23] was Jordan Zlatev's *Semiotic Hierarchy* [36, 35]: a prescriptive account meant to rescue researchers from trapping themselves into implicit assumptions built into the language they have coopted from everyday use. According to the Semiotic Hierarchy, language presupposes (depends upon) sign use, for it seems that language requires the prior ability to communicate. Likewise sign use presupposes a culture of shared intentionality to nurture it, while such culture presupposes some basic level of consciousness: otherwise, one has a society of David Chalmers' philosophical zombies [4]. The intrinsically dynamic nature of consciousness – constantly adapting response to circumstances – in turn presupposes life. Each level (*life - consciousness - culture - semiosis - language*) lays the foundation for the next: a healthy antidote, we thought, to tendencies to overattribute. The definition of creativity in Section 1 is offered with a similarly prescriptive intent.

Whitby and I concluded that moral agency does *not* require linguistic ability [23, p.14];[24, p. 3] but does require semiosis (because the moral agent needs to be able to communicate evidence of her moral agency to herself or others[13]). Similar reasoning should apply to creative agency; and, indeed, Simon Colton and Geraint Wiggins, at least, seem inclined to think that it does, with their talk of creative agents "*framing* their creative acts with information that adds value, possibly via reference to political, historical, or cultural contexts" [9, p. 25]. Nevertheless, language is not required. It seems clear that prelinguistic children of a certain age and mental development can be and are creative in limited but identifiable ways; the same applies, as mentioned earlier, to members of the corvid family. It probably applies to our nearest relatives, the great apes, at least those living in captivity [31]; and it may well apply, in varying degrees, to any number of other species. On the other hand, creative agents presumably *do* need to be able to communicate their creative intent; prelinguistic infants do so to their caregivers, crows to the fellow members of their social group. That is how the creative act is recognized and passed on.

2.2 The signature of life

Setting consciousness aside for the moment, can any existing artefacts claim to be alive? I can think of only two, and they are qualified examples, as only the genome was artefactual (still a significant accomplishment, and a transformational creative act): the recent announcement of a new form

---

[9] Another version of this passage reads, "...it's an aspect of *human* intelligence in general" [1] (emphasis mine).

[10] As Wiggins writes [32, p. 455]: "Self-awareness is generally cited as the property which distinguishes the artist from the craftsperson." Wiggins makes conscious awareness a requirement at least of transformational creativity, if not creativity in general.He discusses the critical role of reflection at much greater length in [33].

[11] This is not to say that artefactual intelligence, conscious or otherwise, can be arrived at in such a way – our stated assumption was and remains that it almost certainly can not – only that, if one *were* presented with an artefactual agent otherwise indistinguishable from a intelligent agent, one should not conclude, as Searle does, that the agent lacks a mind *just because* one "knows" it is only following a rulebook. Either it is not really doing so, despite appearances, or one's assumptions about what makes intelligence possible require revisiting.

[12] Such a definition will, of course, be a nonstarter for those who consider consciousness epiphenomenal and those who follow Daniel Dennett[10] in seeing consciousness as a useful fiction we tell ourselves. It is, however, compatible with Thomas Nagel's [19] "being there"; for how else does one judge whether someone

else has that sense of "being there" – how else does one judge whether anyone is "at home" – except through observation of such behaviour?

[13] Note that this does not require that she actually do so, only that she has the capacity to do so, given the appropriate opportunity.

of the *Escherichia coli* intestinal bacterium based on an entirely synthetically constructed DNA [25] and an earlier effort based on *Mycoplasma mycoides* [13]. Other efforts have mixed pre-existing with synthetically constructed DNA. All of these projects are light years away from producing conscious agents, let alone creative ones. Nonetheless, it seems safe to suppose that creative artefacts of the future will not much resemble the clunky mechanical devices that the term "artefact" often conjures up today, even as they may not bear much resemblance to naturally occurring terrestrial cell-based life, either. Likewise one can expect that their creative expressions will exhibit "*creativity, but not as we know it*" [9, p. 25].

As for what counts as "alive", one should not – on my view, as should be clear when I called the artefactual question a red herring – be prejudiced by what "stuff" something is made of, be it carbon or silicon, cell- or processor-based or something else again. A function-based definition that abstracts away from any familiar laundry list of biological properties seems preferable. I take inspiration from Francisco Varela and Humberto Maturana's notion of *autopoiesis* [16], which deliberately does just that: abstracting away from life-as-we-know-it to describe life in functional terms as a kind of homeostatic machine where "...the organization of a machine is independent of the properties of its components which can be any, and a given machine can be realized in many different manners by many different kinds of components" [16, p. 77].[14] Such a machine defines its own, selectively permeable boundary between itself and the world; is organizationally closed, its structure determined and maintained by processes within itself; and is autonomous in the strong sense, as opposed to giving the partial appearance of autonomy [24, p. 5].

Are human beings – are all organisms on Earth – machines? In a functional sense, yes.

I am not aware of anyone claiming their artefact to be creative who also claims it to be conscious, nor of anyone claiming their artefact to be ("minimally" or otherwise) conscious who also claims it to be alive. Why is that? I think it's because, for many people, creativity seems – on the face of it – like not such a hard think to emulate, even replicate, even formalize. What, they might say, is creativity *but* the artful recombination of past or present experience (whatever exactly "experience" is taken to be)?

Consciousness, on the other hand, seems much harder to get a grasp on – perhaps because conscious agents, seeking to explain consciousness, are necessarily seeking to explain their *own* consciousness, and how does one step outside one's own consciousness to do that? One feels some sympathy for the Churchlands [6, 5] (or, really, Daniel Dennett [10]) who would try to explain and reduce it straight out of existence. What separates life from non-life, organic from inorganic matter, seems that much bigger of a question – and really, 100% synthetic DNA organisms notwithstanding, researchers haven't a clue where to begin – except that putting a bunch of organic

compounds in a bottle and zapping it with energy is probably not going to be sufficient. But creativity, naively perhaps, seems much easier to understand, even encode into a set of rules and random-number generators for a machine. I truly do have sympathy for artist Paul Brown writing [3]:

> A key problem is that of signature: at what point can we claim that an artwork has its own distinct signature?... I suggested that using a symbolic language to initiate a process would distance me far enough from the output of that process for it to have the potential of developing its own intrinsic qualities including a unique signature. By the 1990s it had become obvious that this approach had failed. Complementary research in many fields had demonstrated that the signatures of life were robust and strongly relativistic.

My impression is that the bottom-up "evolutionary" processes employed by the Drawbots project he was involved with at the time did not fare any better than the top-down "symbolic" approach had earlier. Why? Because evolution on its own is no solution if one does not have the right starting point. The signature of life was and is still missing.

2.3 Further requirements

Two other things deserve saying about who qualifies as a creative agent. Both are taken from the proposed list in [24] of who qualifies as a moral agent.

1. *The creative agent must be not just a conceptual agent but a sophisticated one* (*cf.* [24, p. 5]). In order to recognize herself as trying to be creative or trying to express that creativity to others, she must at the least have an explicit concept of self, of other, of creativity, of the creative act itself: concepts that, presumably, not all conceptual agents have. She must possess not just first-order concepts (concepts of that which purports, at least, to be non-concepts) but higher-order concepts (concepts of concepts: most importantly, perhaps, a concept of "I-as-myself" reflecting the ability to have and reflect on her sense of mental presence, of autobiographical narrative). Higher-order concepts turn their attention from the world to themselves – again, an ability that, in all likelihood, not all conceptual agents possess.

2. In parallel fashion, *the creative agent must be not just a conscious agent but an actively self-aware one* (*cf.* [24, p. 6]), for creativity thrives on the capacity for reflection ("what have I done?"). Remember the earlier claim that conceptual agency and consciousness are two ways of looking at the same phenomenon (where one is inclined to attribute the one, one is generally if not highly reliably inclined to attribute the other)? Drawing on a distinction common within phenomenology – between pre-reflective and reflective consciousness (see e.g. [11]) – the creative agent must be not just consciously aware but consciously aware of being consciously aware, where a good starting test for the latter might well be the ability to pass some suitable version of the mirror test.

---

[14] Alan Turing uses "machine" in exactly this sense in [30, p. 435] before then restricting it to "digital computers".

3   Why then computational creativity?

When one's world becomes too small, it is time to break down the walls of that world or push those walls outward [21, p. 77].

The reader might be forgiven for thinking by this point that I find little value in discussions of computational creativity – but that could not be further from the truth. It is true that I see little to gain and much in the way of confusion to lose by attributing creativity to existing artefacts, be they programs floating in a largely virtual world or robots interacting quite directly with ours. I remain agnostic about the possibility of future artefactual creative agents, and I would not even begin to try to guess when they might be achieved: in the next five years, in my lifetime, or not for a thousand years if ever.

However, it was not the promise of independently creative agents that first got me interested in this area, back in my undergraduate days in the early 1980s. It was the potential for human-computer collaboration (see e.g. [9, p. 23]) – though "collaboration" is a loaded term, and one I would just as soon avoid any more – using technology to extend or re-imagine human capacities. It was the promise – but already increasingly then the reality, even then – that humans and computers, working together, could achieve things that lay outside either's reach on their own. Somewhat pretentiously, I called it the "Vegar Effect"[15] in an undergraduate paper, referring to a science fiction film that had come out a few years before.

3.1   Creativity transformed

Already at that time, computers were being used to create mathematical proofs that mathematicians could not arrive at on their own. The computer proof verifying the four-color theorem was achieved in 1976. Already at that time, computers were revolutionizing astronomy, facilitating the collection and processing of data on a scale not previously conceivable. By the late 1960s, computers were being used to produce the first detailed models of what happens when a star goes supernova. By the end of the 1960s, the Internet was becoming a thing, facilitating the exchange of information between researchers and universities in moments where before one had to rely on phone calls or personal meetings at conversation or paper-based publications – remember those? Now, of course, the Internet is a thing on a whole different level – the so-called *Internet of things* – with around 1,000,000 people a day around the world gaining access for the first time. One can pull up one's favourite search engine and have the answer to nearly any question imaginable – if an answer is to be found – in moments; and one has the tools – again, many of them computer-based – to do a pretty good job of evaluating just how reliable one can take the answer to be. All it requires is some skill in critical thinking, which the Internet certainly encourages one to practice. If it's all a bit overwhelming sometimes, it is also breathtakingly creative.

Already at that time, visual artists like Paul Brown were using computers to explore in radically new directions that were not imaginable a few years in the past. If the human artist's

"signature" remained in the digital "artist's" output, what of it? Already at that time, computers were remaking how music got made. They inspired the creation of the first Moog synthesizer in 1964, the year I was born. Before then, synthesizers – like computers of the time – filled entire rooms. By the time I was in junior high school, mini-Moogs were cheap enough and small enough that my school could have one. The same thing, of course, was happening with computers. The first affordable desktop computers started coming out in the mid 1970s. By the early 1980s, they were becoming commonplace on university campuses, and many students had one in their rooms.

Ironically, the name "computer", bestowed on the first digital computers in the 1940s, was a metaphor for human computers: people whose job it was to do calculations, often largely in their heads. Teams of computers would work independently to verify complex calculations. Within a few short years, those teams disappeared. What had been the metaphorical meaning of the word became the "literal" one, and what had been the "literal" meaning became the metaphor: the human mind as computer. Scores of researchers, going back to the earliest years of AI and increasing in numbers over time, took that metaphor as something approaching blasphemy. How, they said, could one compare the messy human mind, with all its creative capacity and all its capacity to get things wrong, to a computer that (many people wrongly thought, and many wrongly still think) can never get a calculation wrong, and can only do what it is told?[16] As Boden writes [2, p. 29]: "Creativity and computers: what could these possibly have to do with one another? 'Nothing', many people would say [and have said, and continue to say, over and over again]. Creativity is a marvel of the human mind" – understood to be more powerful than any formal system.[17] Attacking the metaphor of mind-as-computer became a rallying cry for all who thought there was something fundamentally *wrong* with the project of artificial intelligence. The irony behind the inverted metaphor largely got lost along the way.

3.2   Alan Turing

One person, more than any other, saw the creative potential in digital computers from the very beginning. Indeed, he had a hand in helping build one of the first ones, in Manchester, England. I am thinking, of course, of Alan Turing. Most people who are aware of his seminal paper in *Mind* [30] (which spends much of its time talking about creativity) remember it only for what they call the Turing Test and what Turing himself called the Imitation Game; and a great many of them get the "test" wrong, for Turing never said and almost certainly

---

[15] The correct spelling, according to the film script, would have been "V'ger".

[16] ...A mistaken line of reasoning that can be traced back to Lady Ada Lovelace, for all the value of her other insights. For a good introduction to Lovelace's role in the foundations of modern computing, see [14, p. 25].

[17] The discussion in [26, pp. 72-77] is especially enlightening here. For a discussion of what I think is wrong with Roger Penrose's argument – that, unlike computers, the human mind is *not* bound by Gödel's Incompleteness Theorems – see [20, pp. 192-195]. So far as I can tell, Turing's statement [30, p. 445] that "...although it is established that there are limitations to the powers of any particular [discrete state] machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect" still holds.

did not believe that a computer program, successfully winning at the Imitation Game, could by virtue of having done so be said to have a human or human-like mind: in short, to think as people do. That was not the stated purpose of the game.

The most important point of that paper for me – and, I would like to think, for Turing – was what he saw as the transformative power that digital computers – still in their infancy at the time he wrote the paper – would have in getting the human mind to rethink itself, to rethink the nature of thinking itself. Turing's greatest insight lie in seeing digital computers as a mirror by which the human mind could consider itself in ways that previously were not possible.

Turing made two predictions in the 1950 paper. One of them – that, within fifty years, "the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" [30, p. 442] – has long ago proven its point. He was also, if I am reading him correctly, correct about the creative revolution that digital computers would unleash. What he missed – what he could not help but miss, doubtless could not imagine – was the creative revolution that his paper would unleash, within the AI community and beyond. All its insights aside, his paper is a playful work of creative fancy that has inspired two going on three generations of researchers and is being read by high school students today.

Most interpretations of the 1950 paper are controversial; people will tend to read into it what they seek to find. What is doubtless true is that Turing perceived an underlying mathematical kinship between human mind and digital computer. What is probably true is that he imagined a future in which artefacts of one kind or another – inspired, in their origins, by the digital computers he was helping to create – would take their place as fully fledged citizens in a society where the line between the artefactual and the biologically human would, for many purposes outside procreation, be blurred to the point of meaninglessness. Fanciful? Perhaps; but it does not strike me as logically outside the realms of conceivability.

What does it mean to be human?[18] The power of computational creativity – the coming together of human mind and artefactual entity, a digital calculator that has become so much more – ultimately lies, I suspect, in its ability to force one to confront that question, sometimes in disquieting ways. Given the unprecedented challenges facing *homo sapiens,* the species could use all the creative energy it can get.

---

[18] ... A question I consider at length in [21, p. 81], concluding that "...any strictly biological definition of what it means to be human will be inadequate, even in biology; but even more so in areas outside biology. Why?... because while the boundaries of 'human being' as a biological organism are relatively clear (although ask a biologist, and the biologist will admit that the definition of 'species' is not so clear cut as it might appear at first blush); the boundaries of 'human being' as a cognitive entity are much less so. ..."

References

[1] Margaret Boden. Can computer models help us to understand human creativity? Available from https://nationalhumanitiescenter.org/on-the-human/2010/05/can-computer-models-help-us-to-understand-human-creativity/ (accessed 1 February 2020), 2010.

[2] Margaret A. Boden, *Creativity and art: Three roads to surprise*, Oxford University Press, 2010.

[3] Paul Brown, 'Autonomy, signature and creativity', in *09291 Abstracts Collection: Computational Creativity: An Interdisciplinary Approach (Dagstuhl Seminar)*, (2009). Available from https://drops.dagstuhl.de/opus/volltexte/2009/2204/ (accessed 2 February 2020).

[4] David J. Chalmers, 'Review: Imagination, indexicality, and intensions', *Philosophical and Phenomenological Research*, **68**(1), 182–190, (January 2004). Available online from https://www.jstor.org/stable/40040667?seq=1 (accessed 25 February 2020).

[5] Patricia Smith Churchland, 'Reduction and the neurological basis of consciousness', in *Consciousness in Contemporary Science*, eds., A.J. Marcel and E. Bisiach, 273–304, Clarendon Press, (1988).

[6] Paul M. Churchland, *Matter and Consciousness*, MIT Press, 2013.

[7] Harold Cohen, 'The further exploits of Aaron, painter', *Stanford Humanities Review*, **4**(2), 141–158, (July 1995). Available from http://haroldcohen.com/aaron/publications/furtherexploits.pdf (accessed 25 February 2020).

[8] Simon Colton, 'Creativity versus the perception of creativity in computational systems', in *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*, (2008). Available from https://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-03/SS08-03-003.pdf (accessed 19 February 2020).

[9] Simon Colton and Geraint A. Wiggins, 'Computational creativity: The final frontier?', in *ECAI 12: 20th European Conference on Artificial Intelligence*, eds., L. De Raedt, C. Bessiere, and D. Dubois, volume 242. IOS Press, (2012). doi: 10.3233/978-1-61499-098-7-21.

[10] Daniel Clement Dennett, *Consciousness Explained*, Little, Brown, 1991.

[11] Shaun Gallagher, 'Phenomenological approaches to consciousness', in *The Blackwell Companion to Consciousness*, eds., Max Velmans and Susan Schneider, chapter Phenomenological approaches to consciousness., 686–696, John Wiley, (2007).

[12] Jr. George G. Gallup, 'Chimpanzees: Self-recognition', *Science*, **167**(3914), 86–87, (1970).

[13] Daniel G. Gibson, John I. Glass, Carole Lartigue, Vladimir N. Noskov, Ray-Yuan Chuang, Mikkel A. Algire, Gwynedd A. Benders, Michael G. Montague, Li Ma, Monzia M. Moodie, Chuck Merryman, Sanjay Vashee, Radha Krishnakumar, Nacyra Assad-Garcia, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Lei Young, Zhi-Qing Qi, Thomas H. Segall-Shapiro, Christopher H. Calvey, Prashanth P. Parmar, Clyde A. Hutchison III, Hamilton O. Smith, and J. Craig Venter, 'Creation of a bacterial cell controlled by a chemically synthesized genome', *Science*, **329**(5987), 52–56, (July 2010). DOI: 10.1126/science.1190719.

[14] Douglas Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Penguin, 2000.

[15] Nick Kelly and John Gero, 'Situated interpretation in computational creativity', *Knowledge-Based Systems*, **80**(1), 48–57, (2015). doi: 10.1016/j.knosys.2014.12.005.

[16] Humberto Maturana and Francisco J. Varela, *Autopoiesis and Cognition: The Realization of the Living (Boston Studies in the Philosophy of Science)*, Kluwer Academic Publishers, 1980.

[17] Max-Planck-Gesellschaft. New Caledonian crows can create compound tools. Available from https://www.sciencedaily.com/releases/2018/10/181024 112201.htm (accessed 1 February 2020), October 2018.

[18] *The New Oxford Annotated Bible: Revised Standard Version*, eds., Herbert G. May and Bruce G. Metzger, Oxford University Press, 1973.

[19] Thomas Nagel, 'What is it like to be a bat?', *The Philosophical Review*, **83**(4), 435–450, (October 1974). doi: 10.2307/2183914.

[20] Joel Parthemore, *Concepts Enacted: Confronting the Obstacles and Paradoxes Inherent in Pursuing a Scientific Understanding of the Building Blocks of Human Thought*, University of Sussex, 2011. DPhil thesis available from http://sro.sussex.ac.uk/id/eprint/6954/ (accessed 2 February 2020).

[21] Joel Parthemore, 'The "final frontier" as metaphor for mind: Opportunities to radically re-conceptualize what it means to be human', in *The History and Philosophy of Astrobiology: Perspectives on Extraterrestrial Life and the Human Mind*, chapter 3, 67–92, Cambridge Scholars, (2013).

[22] Joel Parthemore, 'Consciousness, semiosis, and the unbinding problem', *Language & Cognition*, **54**, 36–46, (May 2017).

[23] Joel Parthemore and Blay Whitby, 'What makes any agent a moral agent? reflections on machine consciousness and moral agency', *International Journal of Machine Consciousness*, **5**(2), 1–25, (2013).

[24] Joel Parthemore and Blay Whitby, 'Moral agency, moral responsibility, and artifacts: What existing artifacts fail to achieve (and why), and why they, nevertheless, can (and do!) make moral claims upon us', *International Journal of Machine Consciousness*, **6**(2), 1–21, (2014).

[25] Roland Pease, 'Artificial life form given "synthetic DNA"', *BBC News*, (May 2019). https://www.bbc.com/news/science-environment-48297647 (accessed 2 February 2020).

[26] Roger Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford University Press, 1994.

[27] Uma Ramamurthy and Stan Franklin, 'Self-system in a model of cognition', in *Machine Consciousness 2011: Self, Integration and Explanation, Proceedings of AISB'11, 4-7 April 2011, York, UK*, eds., Robert Clowes, Steve Torrance, and Ron Chrisley, pp. 51–54. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), (2011).

[28] John Searle, 'Minds, brains, and programs', *Behavioral and Brain Sciences*, **3**(3), 417–458, (1980).

[29] Zoltan Torey, *The Crucible of Consciousness: An Integrated Theory of Mind and Brain*, MIT Press, 2009.

[30] Alan Turing, 'Computing machinery and intelligence', *Mind*, **59**(236), 433–460, (1950). Available online from many sources including https://www.abelard.org/turpap/turpap.php (accessed 2 February 2020).

[31] C.P. van Schaik, J. Burkart, L. Damerius, S.I.F. Forss, K. Koops, M.A. van Noordwijk, and C. Schuppli, 'The reluctant innovator: orangutans and the phylogeny of creativity', *Philosophical Transactions of the Royal Society B*, **371**(1690), (2016). DOI: 10.1098/rstb.2015.0183.

[32] Geraint A. Wiggins, 'A preliminary framework for description, analysis and comparison of creative systems', *Knowledge-Based Systems*, **19**(7), 448–458, (November 2006). doi:10.1016/j.knosys.2006.04.009.

[33] Geraint A. Wiggins, 'Searching for computational creativity', *New Generation Computing*, **24**, 209–222, (2006).

[34] Jordan Zlatev, 'Meaning = life (+ culture): An outline of a unified biocultural theory of meaning', *Evolution of Communication*, **4**(2), 253–296, (January 2000).

[35] Jordan Zlatev, 'The semiotic hierarchy: Life, consciousness, signs and language', *Cognitive Semiotics*, **4**, 169–200, (2009). Available from https://www.academia.edu/download/49336286/Zlatev.CS.2009.pdf (accessed 1 February 2020).

[36] Jordan Zlatev, 'Meaning making from life to language: The semiotic hierarchy and phenomenology', *Cognitive Semiotics*, **11**(1), (2018). doi: 10.1515/cogsem-2018-0001.

33

AMI – Creating Musical Compositions with a Coherent Long-term Structure

Ning Ma and **Guy J. Brown** and **Paolo Vecchiotti** [1]

Abstract.

We present AMI – Artificial Music Intelligence, a deep neural network that can generate musical compositions of different instruments with a coherent long-term structure. AMI uses a state-of-the-art general-purpose deep neural network architecture, called the Transformer model [7], to discover patterns of musical structures such as melodies, chords, and rhythm, from tens of thousands of MIDI files. The learning is done in an unsupervised manner, allowing exploitation of large collections of MIDI files that are available on the internet. We trained AMI over 8000 classical music MIDI files. As an autoregressive model, AMI predicts a music note at a time depending on not just the last note, but a long sequence of notes (up to thousands) from previous time steps. The previous notes are not provided via some hidden state such as in a recurrent neural network (RNN), instead the model has direct access to all earlier notes. Furthermore, we enhance the learning of musical structures by adding different kinds of embeddings: one short-term embedding and one long-term embedding. As a result, the model is able to maintain a coherent long-term structure and occasionally pick up different movements. Audio examples of the model output can be heard at `https://meddis.dcs.shef.ac.uk/melody/samples`.

The Transformer model is the latest advance in language understanding [6, 2, 1], which is trained to predict the next token in a sequence of text. The core idea behind the model is self-attention – the ability to attend to different positions of the sequence to compute a representation of that sequence. It allows the modelling of a much longer sequence than one that can be modelled by previous language models such as a recurrent neural network based model. This makes the model well-suited for modelling music data, whose structure and meaning are often built by repetition and self-reference on multiple timescales.

Unlike one-dimensional text data, music has multiple dimensions; the two most prominent are timing and pitch. To model musical data using a language modelling approach, we first convert musical data to a sequence of discrete tokens. One popular representation was proposed by Oore et al [5], who encoded musical data using an event-based representation. Similar to MIDI events, the encoding consists of a vocabulary of 128 NOTE ON events, which correspond to possible MIDI pitches in the range [0, 127], 128 NOTE OFF events, 100 TIME SHIFT events and 32 VELOCITY events for modelling expressive dynamics. A sequence of consecutive note events between two TIME SHIFT events are considered to be played simultaneously. The 100 TIME SHIFT events span a duration of one second, allowing for expressive timing at 10 ms. This representation was also used in the Music Transformer model proposed by the Magenta team at

Google [4]. One of the drawbacks of this representation is that it encodes music in time units. As a result, the tempo information is implicitly encoded by the TIME SHIFT events which has a resolution of 10 ms. For slow tempo music this may be oversampling which generates redundant event tokens but may become undersampled for fast tempo music. More importantly, representing music in time units has no explicit representation of musical notations which makes it more difficult to encode music structures.

We propose an encoding method that encodes MIDI files in units of beats (quarter notes). We sample each beat 12 times, which allows us to also handle a range of time signatures. Instead of encoding note on/off events, each note is directly represented by its pitch and its duration in the unit of 1/12 beat. Such an encoding method has a closer resemblance to a music score, which shows note pitches as well as note values. As now note durations are represented in beats, we also explicitly encode the change of tempo events. This allows tempo changes in the middle of a music piece without changing note durations similar to a music score representation. Following the Oore [5] encoding, velocity events are also encoded for modelling expressive dynamics. Finally, we encode instrumentation tokens to represent the instruments that should be used to play the notes that follow an instrumentation token. Instrumentation tokens are separated from note tokens to reduce the number of tokens in the model vocabulary, but this also allows the model to learn chords and melodic relations from more data and have the freedom to swap instruments during generation in order to become more creative. Currently we encode the following 8 instruments:

```
<drum> <piano> <guitar> <bass> <strings>
<brass> <reed> <pipe>
```

An example music score is given in Fig. 1. Here the tempo is 80 BPM and the MIDI file is played by piano. Each note is presented by a combination of its pitch and its duration, e.g. A2:24 is piano key A2 playing a minim (two beats, or 24 time steps at a sampling rate of 12 time steps per beat). For notes that are played simultaneously, e.g. a chord, we sort them by pitch from low to high. The "wait" tokens encode how long (in time steps) it needs to wait before next events. This encoding method is akin to reading a music sheet.

For the current version of AMI, we trained a 12-layer transformer model with 12 attention multiheads [7] – with full attention over a

[1] Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK, email: {n.ma, g.j.brown, p.vecchiotti}@sheffield.ac.uk

```
tempo:80 <piano> vel:88 A2:24 E3:24 C4:12
wait:12 D4:12
wait:12 C3:24 G3:24 E4:24
<EOS>
```

Figure 1. An example music score and its encoded tokens. EOS represents end of music score.

context of 1024 tokens. This is roughly equivalent to 400 beats with the proposed encoding methods. The dimension of the token embeddings is 768. To enhance learning of the musical structure, we added different kinds of embeddings. The first embedding is short-term and encodes the beat information so that all the tokens belonging to the same beat are given the same beat embedding. This helps the model to keep a consistent rhythm throughout a piece. The second one is a long-term structural embedding and encodes musical phrases. In the current version 8 measures are considered to form a musical phrase and all the notes that form a musical phrase are given the same phrase embedding. This gives the model more information about the overall structure of a musical piece.

We collected training data for AMI from different on-line sources, including 4000 MIDI files from `https://www.classicalguitarmidi.com` and the MAESTRO dataset [3] which contains over 200 hours of virtuoso piano performances. To avoid overfitting and help the model generalise key scales, we further augmented the training data by transposing each MIDI file to different keys.

So far AMI has been trained mainly on classical music and we plan to train it on more music genres. A couple of example outputs are given below. For the first example shown in Fig.2, AMI was asked to generate music from scratch, given a key signature G Major and an initial tempo of 100 BPM. The instrumentation and the arrangement were both selected automatically by AMI. Guitar was selected to accompany flute melodies and the two instruments interplayed in harmony throughout the generated piece. AMI was also able to maintain a coherent long-term structure. It is also interesting to note that the model generated almost different movements, e.g. from the 9th measure.
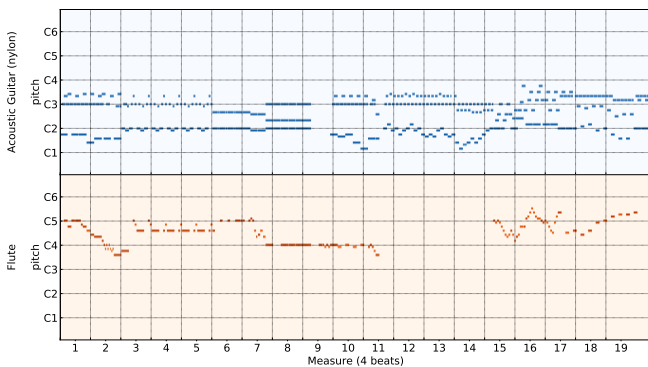

Figure 2. An example output from the AMI. The key signature G Major and an initial tempo of 100 were specified. Both the instrumentation and the arrangement were selected automatically by the model.

In the second example shown in Fig.3, AMI was prompted by 8 beats from Chopin's Nocturne Op.9 No.2 and asked to continue for 120 beats. It is interesting to notice that some melody snippets are repeated and there is a level of self-reference on multiple timescales. The left hand chords that form the 3-beat rhythm, a theme throughout Chopin's Nocturne op.9 No.2, can be heard throughout the generated
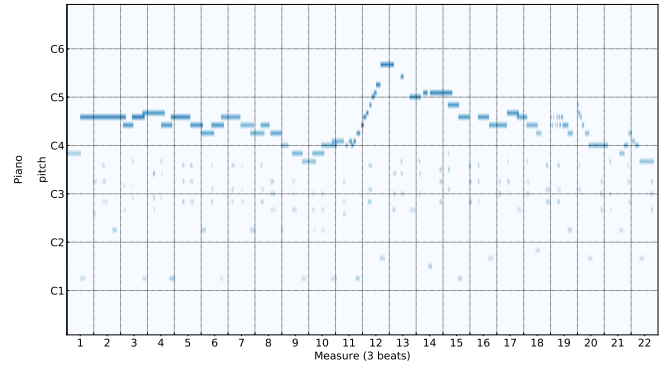

Figure 3. An example of generation prompted by 8 beats from Chopin's Nocturne Op.9 No.2. AMI was asked to continue the melodies. The 3-beat rhythm, a theme throughout Chopin's Nocturne op.9 No.2, can be heard throughout the generated piece.

piece. It is also noticeable that the generated accompanying chords for left hand are consistently quieter than the right hand melodies.

The audios of the two examples presented here and more audio examples can be heard at this link below [2].

ACKNOWLEDGEMENTS

REFERENCES

[1] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G. Carbonell, Quoc V. Le, and Ruslan Salakhutdinov, 'Transformer-xl: Attentive language models beyond a fixed-length context', *CoRR*, **abs/1901.02860**, (2019).

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.

[3] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the maestro dataset, 2018.

[4] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew Dai, Matt Hoffman, Monica Dinculescu, and Douglas Eck, 'Music transformer: Generating music with long-term structure', in *Proc. ICLR 2019*, (2019).

[5] Sageev Oore, Ian Simon, Sander Dieleman, Douglas Eck, and Karen Simonyan, 'This time with feeling: learning expressive musical performance', *Neural Computing and Applications*, (2018).

[6] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever, 'Improving language understanding by generative pre-training', (2018).

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin, 'Attention is all you need', in *Advances in Neural Information Processing Systems 30*, 5998–6008, (2017).

---

[2] `https://meddis.dcs.shef.ac.uk/melody/samples`

Jazzy Beach Critters: a Demonstration of Real-Time Music Generation with Application to Games

Donya Quick[1] and Christopher N. Burrows[2]

Abstract. We present a system design for applying functional models of improvisational music to the generation of game soundtracks that react to game events. We provide a proof-of-concept implementation using Unity/C# and Haskell to create an interactive scene where the music changes as the user interacts with creatures in the environment.

1 Introduction

Procedural generation of various aspects of games is becoming increasingly common. Maps, locations of items, and characters are all frequent points of randomization. There has also been significant work in the area of procedural music generation for games. However, highly procedural music generation of game soundtracks is not as frequent an occurrence as is, for example, procedural map generation. There has been some work on fully procedurally generated scores for game soundtracks[1, 2, 4], but most mainstream games that attempt to adapt the soundtrack to game events do so by cross-fading between existing recordings. There is still relatively little work on novel score generation for games focused on note-to-note generation instead of heavy reliance on recombinance of existing musical material (whether at the score or audio level).

In a real-time generative scenario where user-triggered events occur unexpectedly (as contrasted with movies or cutscenes where the timing is known in advance), there are fundamentally two kinds of procedural sound: instantaneous sounds or sound effects, for which immediate playback is most important, and music, for which maintaining hierarchical consistency can be more important than providing an immediate response. Music can become disorganized if changed too quickly, and as such may not be able to exactly synchronize with user-triggered events while still maintaining coherency. Cross-fading rapidly between a slow, steady piece of music and a fast, syncopated one to try to match a change in mood is an example of this. If handled poorly, the transition will be both noticeable and potentially distracting from gameplay.

Markov chains are one of the few strategies that are widely used in music while also having been applied to note-by-note sountrack generation in games[3]. While Markov chains can adapt to game changes both quickly and smoothly, they are prone to state space explosion when modeling complex structures. Hierarchical Markov chains have been used in game music generation as a means to mitigate this while retaining some musical structure[1]. Even so, these models are not

---

[1] Stevens Institute of Technology, email: donyaquick@gmail.com
[2] Texas A&M University, email: christopher.n.burrows@gmail.com
[3] For a broader overview of note-level generative strategies for music (not necessarily in the area of games), we refer the reader to our prior work[3], which describes our generative music strategy in more detail and makes comparisons to other methods.

Figure 1: Screenshot of "Jazzy Beach Critters" showing the three critters: a hermit crab (left), snail (middle) and true crab (right). Randomly colored note symbols are emitted like particles by the critters each time they play a musical note.

well-suited to situations like ours where each part is best treated as a partially-independent process. In our generative model, each instrument has its own independent internal state and, at each generative call, each part can generate multiple notes (whether sequentially as a melody or simultaneously as a chord).

We have created an interactive scene, called "Jazzy Beach Critters," which allows the preservation of larger-scale structures in real-time generated music while still adapting to changes in style and mood based on user actions. The scene features a beach at sunset where the user can interact with animals as music is created. Much like the famous "Peter and the Wolf," the animals in our scene correspond to instruments, and the music they generate reflects their mood. The critters collectively make up a jazz band with the rolls of lead, harmony, and bass. The ocean is a fourth participant in the music and plays a drum beat that is synchronized to the appearance of waves. Notes emitted by the critters (and ocean) must be synchronized to share a harmonic context, metrical structure, and given tempo. The harmonic context is randomized and can change periodically. A video demo of the scene is available online at:

`www.donyaquick.com/jazzy-beach-critters`

The rate at which the structured music can be changed depends on two things: the complexity of the generative algorithm and the minimum temporal span of important features. In our case, a single measure of music playing out before changes take effect was sufficient to accommodate both of these limitations with a 4/4 time signature even at elevated tempos. Practically, this means that the music will only take 1-2 seconds to adapt while maintaining harmonic and metrical coherency.

Our generative model for music is taken from "A Functional Model of Jazz Improvisation"[3], which presents a strategy for generating improvisational music in real time while communicating information between musicians and adhering to shared constraints (such as the current key). Each part or instrument stochastically generates short segments of music while maintaining an internal state, which is updated between segments. In a real-time setting, generation of segments can be overlapped slightly to compensate for generation time and ensure a continual stream of music. The generative workflow we used is shown at a high level in Figure 2. We emphasize two points: (1) communication between between the generative music algorithms and game scene is *bidirectional*, allowing music to influence game events in addition to game events influencing the music, and (2) our generative models are not reliant on recombinance of existing music and can produce completely novel sequences.

2 Implementation

Each critter's sounds are spatial such that the critter's location in the scene affects panning. Each critter has three moods that govern its musical behavior:

- *Happy* critters play cohesively in the current style and obey the shared harmonic and metrical context.
- *Neutral* critters play only a few notes periodically but still adhere to the shared harmonic and metrical context.
- *Angry* critters follow the metrical context but now disregard the harmonic context and produce dissonance. If all three critters are angry, the result will be largely atonal.

Users can interact with the critters and scene in several ways:

- *Placing food* by clicking on the ground. Nearby critters will move to eat the food, allowing users to move critters around the scene. The food will eventually disappear if not consumed.
- *Petting a critter*, which makes the critter move closer to the user and improves the critter's mood if neutral or angry.
- *Poking a critter*, which makes the critter run away; its mood will sour if happy or neutral.
- *Changing the style* for happy critters by clicking a beach ball in the scene.

"Jazzy Beach Critters" is implemented with a combination of Unity and Haskell. Graphical elements and determination of high-level musical features like style and critter mood take place within Unity and C# (Unity's scripting language). The music generation is implemented in Haskell[4] as a script that takes arguments for each critter and returns a score represented as abstract notes (pitch and onset for each sound).

In order to better synchronize audio and visual elements, sound synthesis from the abstract notes takes place back in Unity/C#. Each critter has multiple channels for triggering audio samples in round-robin fashion to allow notes to overlap in time. The music generation script is called partway through the playback of the last musical segment and runs asynchronously to return the next musical segment before the current one ends. In the unlikely event that the script is unable to finish in time, the current segment is simply repeated to ensure a continuous stream of music. In addition to structured music, "Jazzy Beach Critters" also features sound effects corresponding
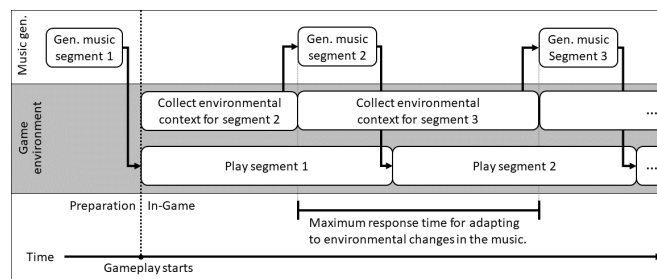
Figure 2: A relative depiction of our workflow over time for generating and playing segments of music. Segments typically contain many notes. In "Jazzy Beach Critters," each segment spans one measure of music.

to user and critter actions. These events are unlikely to align with the music and simply serve as auditory feedback when actions are performed.

Scheduling is handled differently for sound effects and structured musical sounds. Sound effects can be simply triggered in the graphical update loop while achieving a satisfactory result, but this does *not* work for structured musical elements where consistent spacing of the events in time is actually more important. Triggering musical note playback at the graphical update rate results in audible jitter in the metrical structure, and so the samples must be triggered using Unity's separate audio timing system.

3 Conclusion and Future Work

"Jazzy Beach Critters" is a proof of concept for using real-time models of improvisational music in a game scene where features of the music need to change in response to user-triggered events. While this particular demo is designed to facilitate control over events that change the music, addition of more classic game features like goals and challenges would be simple - for example, allowing the critters to affect each others' moods, thereby making it trickier to achieve cohesive music for an extended period of time. We feel that the same system design can be extended into more complete, larger games to create a unique musical experience for each play through and to add depth to the gameplay experience.

Currently, our critters only play jazz, but the same musical framework we have used here has also been applied to other genres and has even been used in live, interactive performances with human musicians. Introduction of other styles into our scene would be easy, as would the introduction of musically-meaningful interactions such as the incorporation of motifs provided by the user into the critters' output. Style could also be changed on a per-critter basis rather than as a group, as our generative music framework can be used to mix-and-match parts of different styles.

REFERENCES

[1] Steve Engles, Fabian Chan, and Tiffany Tong, 'Automatic real-time music generation for games', in *Proceedings of the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, pp. 220–222, (2015).
[2] David Plans and Davide Morelli, 'Experience-driven procedural music generation for games', **4**, 192–198, (2012).
[3] Donya Quick and Kelland Thomas, 'A functional model of jazz improvisation', in *Proceedings of the 7th ACM SIGPLAN International Workshop on Functional Art, Music, Modeling, and Design*, p. 11–21, (2019).
[4] Judy Robertson, Andrew de Quincey, Tom Stapleford, and Geraint Wiggins, 'Real-time music generation for a virtual environment', in *ECAI 98 Workshop on AI/ALIFE and Entertainment*, (1998).

---

[4] The decision to use Haskell is purely a language preference on the part of the authors due to the functional nature of some of the musical models. It would be entirely possible to implement the same algorithms in C#.

Walk the Line:
Digital Storytelling as Embodied Spatial Performance

Philipp Wicke [1] and **Tony Veale** [2]

Abstract. *Demonstration Abstract for the Show-and-Tell session*

We present a robotic storytelling system, named *Scéalability*, which augments a symbolic story-generation system with embodied, robot actors to physically enact a story with the congruent use of space, gesture and voice. We describe the system and summarize the empirical evidence as to the benefits of embodied story-telling.

1 IN THE LIMELIGHT

The generation of plots, stories, poems and literary artifacts has come a long way in Computational Creativity, from early symbolic attempts such as Tale-Spin [5] to the most recent trends in machine learning, such as Transformer models [7]. On one hand, research on story generation focuses on creativity [3], socio-cultural semantics [8], suspense [2], interaction [12], and other features of the generated text. On the other, stories are made to be told. Storytelling happens in the physical world, and exploits the affordances of the tellers. Human story-tellers use more than their voices. They put their backs into a good story. We believe machines should do the same.

Robotic storytellers that use apt gestures and meaningful movements are more expressive [1]. Most robotic storytellers use existing stories in which the movements and gestures are "baked into" the script. We focus here on robotic enactment of stories that the machine invents for itself, and for which its actors choose their own gestures and make their own use of the stage. Gestures are iconic (often pantomimic), such as when one robot goes down on one knee to propose to another. These are as much idioms as the turns of phrase in the story's text and spoken dialogue. But our robot actors also space in accordance with cognitive-linguistic accounts of image-schematic reasoning [4]. They move closer when the story fosters emotional connections, and move apart when relations cool. Fig. 1 shows two robots enacting a story in which A proposes to B. The robot on the left (A) performs a proposal gesture; if the robot on the right (B) accepts, it will move closer. If B declines, it will take a step back.

Scéalexability is an embodied, multi-agent system that builds in the *Scéalextric* [9] story-generation system. Since the latter is a modular, symbolic system, it exposes hooks at every level of a story for a performing system to exploit. In this case, our troupe of performers includes two NAO robots, who play the roles of story characters, and an Amazon Echo with Alexa, who assumes the duties of narrator.

---
[1] University College Dublin, Ireland, email: philipp.wicke@ucdconnect.ie
[2] University College Dublin, Ireland, email: tony.veale@ucd.ie

Figure 1. Scene from a performance. The left robot (A) proposes to robot (B). In the background, an Amazon *Alexa* device narrates the action.

2 THE PLOT THICKENS

In the simplest case, Alexa narrates the story as written by *Scéalextric*, and the two robots enact the roles of characters A and B (whomever they may be in the story) with apt physical gestures and spatial movements. However, *Scéalexability* adds both a dialogue model and a dialogue meta-model to *Scéalextric*. The dialogue model gives the robots something to say as they perform their embodied actions (see the subtitles in Fig. 1). The dialogue *meta*-model is needed to give the robots something to say and do when the story focalizes characters other than A and B, such as A-spouse and B-friend. With only two robots, only two characters can be embodied in the performance. All others are alluded to, referenced but not seen.

A series of user studies will allow us to assess the value of pantomime and image-schemas in telling a story. Wicke *et al.* [12] studied the use of gesture by a single robot, while Veale *et al.* [11] added Alexa to the mix to create a story-telling *Double Act*. Those systems allowed the robot's use of gesture to be evaluated, but it takes two robots to evaluate the relative use of space. Unlike previous authors, we can use *Scéalexability* performances to assess the extent to which gestures contribute to a story, as well as the impact of schematic spatial movements. Which adds most to the telling? Do audiences appreciate stories more when gestures are apt, or is it enough that they are eye-catching? Should actors use space congruently, as image-schema theory predicts [4, 10], or do audiences not register the extra effort?

Storytelling performances benefit from embodied tellers whose presence can be felt. A single embodied teller is a one-man-band, and no clear distinction can be drawn between teller and character, model and meta-model. Multiple embodied agents allow devices to

become characters, and to use space and gesture to signal their emotional stances to one another. Their physical space becomes a stage in which a story becomes theatre.

3 GETTING THE SHOW ON THE ROAD

What follows is an extract of a *Scéalextric* story that has been transformed into a stage play by *Scéalability*. N denotes the narrator, A and B the two main characters which will be embodied by two NAOs.

A=Boris Johnson; **B**=Teresa May; **A-enemy**=Jeremy Corbyn; **N**=Narrator

N: What if Boris Johnson fell in love with Teresa May? Something clicked in Boris when confident Teresa came along.
A:[fall_in_love_with]{closer} ”You are as appealing as a monster truck.”
B: ”Why thank you.”
N: So at first, it was said that Jeremy Corbyn had never loved stern Teresa as much as Boris loved her now.
... (redacted) ...
A:[study_under] ”Please take me as your student.”
B:{lecture} ”I am happy to be your guide.”
N: So Boris became a student of Teresa in the ways of promoting conservative values.

As the performance begins, the robot actors introduce themselves as the characters they will portray. In this play, *Jeremy Corbyn* is spoken of but neither seen nor heard. Only *Boris* (A) and *Teresa* (B) are directly embodied on the stage. Notice how the *Scéalability* script provides stage directions to the actors (e.g. as Boris compliments Teresa, robot A is directed to move closer to robot B). But the stage script does not contain explicit gestures. Rather, the robots decide what gestures to perform on the basis of the actions in the script (such as *fall_in_love_with*). The dialogue model associates stock responses with the actions of the *Scéalextric* story system, which entails providing pre-fabricated dialogue snippets for A and B for over 800 story verbs. But the dialogue model is also capable of ad-libbing within certain structural guidelines. For instance, it generates the simile "as appealing as a Monster truck" for Boris to use of Teresa on the fly,

There are two dimensions of physical embodiment at the core of this research. One is the mapping of iconic gestures to story verbs. For example, Fig. 1 shows robot A performing the iconic "bending of the knee" associated with the action *propose to*. We have empirically validated that coherent gesture mappings are appreciated over incoherent ones [under review]. Another is the spatial movement between robots, which maps character dynamics into physical space. Each robot knows its position relative to centre-stage, and can move 3 steps back or forth. Each coordinates with the other to avoid collisions (see Fig. 2). To prevent accidents, subtle gestures (with no wild swings) are preferred when the robots are within 3 steps of each other. Empirical studies show that coherent, schematic uses of space add as much as dramatic gestures to the appreciation of a story, and significantly more than random or incoherent uses [under review].

4 THE CURTAIN CALL

In our studies to date, we have dissected robotic performances and weighed iconic gestures against schematic spatial movements. The use of space here is inherently metaphorical, but not in any showy sense. Robots, like us, have an embodied presence in physical space,
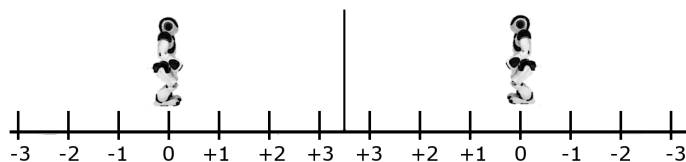

Figure 2. Schematic depiction of the spatial map that both robots use to signify their character relations. The robots will not cross the "+3" line in order to avoid bumping into one other. Moreover, their gestures become more subtle when the distance between robots is less than 3 units.

and their actions should tap into our deep-seated intuitions about the meaning of space [4]. For their part, non-pantomimic gestures are also most effective when they instantiate underlying image schemas that give them metaphorical resonances [6].

By demonstrating the *Scéalability* framework in a Show-and-Tell session, we can elicit feedback and criticisms that can be incorporated into subsequent evaluations of the system. A video of the spatial movement condition can be viewed at `https://youtu.be/RpS8Asl4zjI` and a video of gestural pantomime condition can be viewed here: `https://youtu.be/_IjOjmfoSKo`. The show-and-tell demonstration will integrate both conditions, to work hand in hand on a range of novel, computer-generated tales.

To conclude, this research integrates various strands of research on story generation, story-telling, robot embodiment, and the cognitive linguistics of gesture and spatial schemas. We see it as an initial but promising foray into the realm of embodied computational creativity that views stories (and related products) as more than textual artifacts.

REFERENCES

[1] Adam Csapo, Emer Gilmartin, Jonathan Grizou, JingGuang Han, Raveesh Meena, Dimitra Anastasiou, Kristiina Jokinen, and Graham Wilcock, 'Multimodal conversational interaction with a humanoid robot', in *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 667–672. IEEE, (2012).
[2] Richard Doust and Pablo Gervás, 'Content determination for chess as a source for suspenseful narratives', in *Proceedings of the 3rd Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2018)*, pp. 26–33, (2018).
[3] Pablo Gervás, 'Computational approaches to storytelling and creativity', *AI Magazine*, **30**(3), 49–49, (2009).
[4] Mark Johnson, *The body in the mind: The bodily basis of meaning, imagination, and reason*, University of Chicago Press, 1987.
[5] James R Meehan, 'Tale-spin, an interactive program that writes stories.', in *Ijcai*, volume 77, pp. 91–98, (1977).
[6] Irene Mittelberg, 'Gestures as image schemas and force gestalts: A dynamic systems approach augmented with motion-capture data analyses', *Cognitive Semiotics*, **11**(1), (2018).
[7] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever, 'Language models are unsupervised multitask learners', *OpenAI Blog*, **1**(8), (2019).
[8] Iván Guerrero Román and Rafael Pérez y Pérez, 'Social mexica: A computer model for social norms in narratives.', in *ICCC*, pp. 192–200, (2014).
[9] Tony Veale, 'Déja vu all over again', in *Proceedings of the International Conference on Computational Creativity*, (2017).
[10] Tony Veale and Mark T Keane, 'Conceptual scaffolding: A spatially founded meaning representation for metaphor comprehension', *Computational Intelligence*, **8**(3), 494–519, (1992).
[11] Tony Veale, Thomas Mildner, and Philipp Wicke, 'Duets ex machina: On the performative aspects of "double acts" in computational creativity', in *Proceedings of the International Conference on Computational Creativity*, (2018).
[12] Philipp Wicke and Tony Veale, 'Interview with the robot: Question-guided collaboration in a storytelling system.', in *ICCC*, pp. 56–63, (2018).