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.

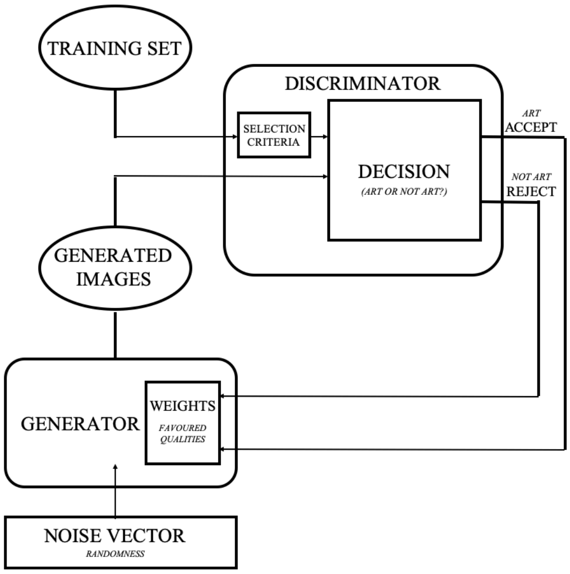


Figure 1 Generative Adversarial Network. Based on information from Goodfellow, I.J. et al. [13]

5 THE CREATIVE ADVERSARIAL NETWORK

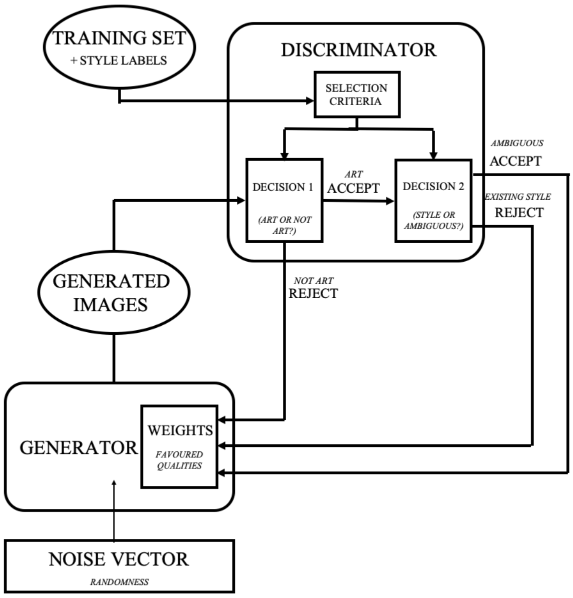
The Creative Adversarial Network (CAN) works with the same basic premise as GANs. It consists of a generator and a discriminator [14]. The training set is also composed of images, however there is the addition of style labels (in Elgammal et al.'s original model, these were artistic style labels, such 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).

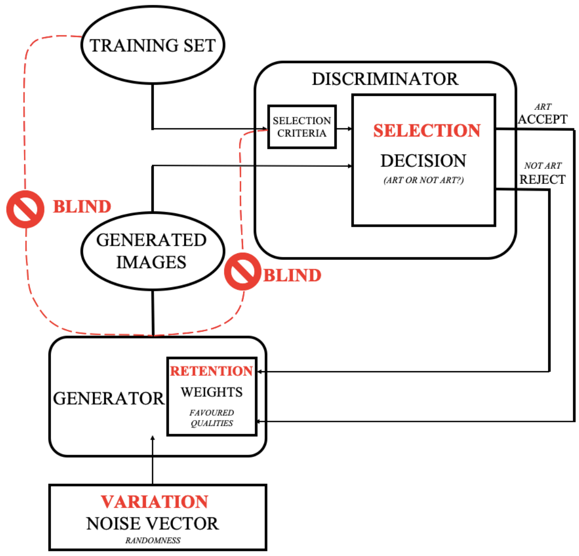


Figure 3 Darwinian model applied to GANs

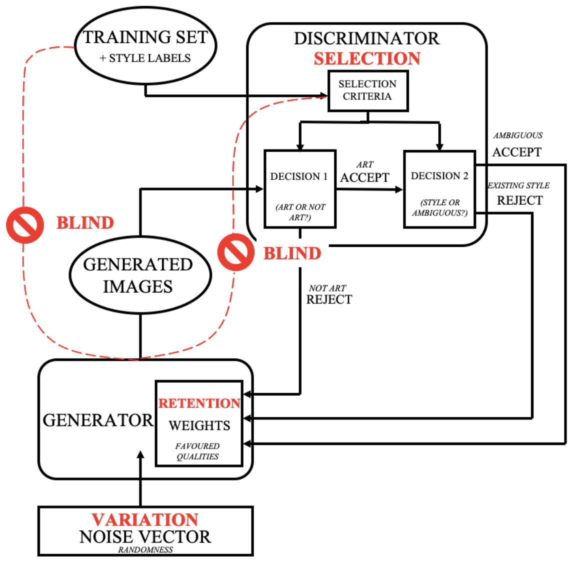


Figure 4 Darwinian model applied to CAN

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.