DOI: http://dx.doi.org/10.7551/978-0-262-31050-5-ch011

Beyond Open-endedness: Quantifying Impressiveness

Joel Lehman and Kenneth O. Stanley

University of Central Florida, Orlando, FL 32826
jlehman@eecs.ucf.edu, kstanley@eecs.ucf.edu

Abstract

This paper seeks to illuminate and quantify a feature of natural evolution that correlates to our sense of its intuitive greatness: Natural evolution evolves *impressive* artifacts. Within artificial life, abstractions aiming to capture what makes natural evolution so powerful often focus on the idea of *open-endedness*, which relates to boundless diversity, complexity, or adaptation. However, creative systems that have passed tests of open-endedness raise the possibility that open-endedness does not always correlate to impressiveness in artificial life simulations. In other words, while natural evolution is both open-ended and demonstrates a drive towards evolving impressive artifacts, it may be a mistake to assume the two properties are always linked. Thus to begin to investigate impressiveness independently in artificial systems, a novel definition is proposed: Impressive artifacts readily exhibit significant design *effort*. That is, the difficulty of creating them is easy to recognize. Two heuristics, rarity and re-creation effort, are derived from this definition and applied to the products of an open-ended image evolution system. An important result is that that the heuristics intuitively separate different reward schemes and provide evidence for why each evolved picture is or is not impressive. The conclusion is that impressiveness may help to distinguish open-ended systems and their products, and potentially untangles an aspect of natural evolution's mystique that is masked by its co-occurrence with open-endedness.

Introduction

A significant challenge in artificial life is to create an evolutionary system with dynamics and products similar in spirit to those of natural evolution. Some researchers believe that a truly *open-ended* evolutionary system will be a critical step towards that goal (Bedau et al., 1998; Standish, 2003). Though the definition of such open-endedness is still debated (Bedau et al., 1998; Lehman and Stanley, 2011a; Maley, 1999; Standish, 2003), there are a variety of reasonable intuitions about what constitutes open-endedness, e.g. increasing complexity, diversity, accumulation of novelty, and continual adaptation. Such intuitions typically are inferred from widely-accepted examples of open-ended evolution like natural evolution or the evolution of technology.

Some have attempted to quantify these intuitions (Bedau et al., 1998; Standish, 2003). Evolutionary activity statistics (Bedau et al., 1998) are the most popular of such measures, and have been applied to many artificial life simulations (Bedau et al., 1997, 1998; Channon, 2001; Maley, 1999; Taylor and Hallam, 1998). The main idea motivating activity statistics is that an unboundedly open-ended evolutionary system will continually accumulate and preserve new adaptations. However, while several systems have passed the test (Channon, 2001; Maley, 1999), they do not seem to meet the high standard set by evolution in nature. The problem is that while the test indicates that adaptations accumulate, it does not reveal their *purpose*. As a result, it is difficult to decide whether the products of such systems are increasingly *impressive* (Channon and Damper, 2000; Maley, 1999). In other words, an increasing diversity of adaptations may not be a sufficient condition for what we appreciate intuitively about natural evolution. This possibility hints that open-endedness and impressiveness may not always be linked.

Approaching intuitions about evolution from a different perspective, this paper argues that a key feature of impressive open-ended systems like natural evolution is that their *products* are indeed impressive. For example, consider the human brain or the wide variety of complex animals crafted by natural evolution. Among their many features, they are usually regarded as impressive achievements. Yet what does impressiveness actually mean? Well-adapted natural organisms, elegant technological innovations, masterful human paintings, and great musical compositions all share the property that they are *easier to appreciate than to create*. Similarly to the concept of NP-completeness, wherein a computational solution is easy to verify but difficult to derive, this paper posits that impressive artifacts are those that readily exhibit significant design effort. In other words, it is easy to appreciate for an impressive creation how difficult recreating an artifact with similar properties would be.

This new formalization leads to two heuristics for quantifying the impressiveness of evolved products, *rarity* and *re-creation effort*, which are applied in this paper to an exploration-driven picture-evolution system. The results establish that the system discovers increasingly impressive artifacts compared to a random search or a direct search for rare artifacts. Importantly, what in particular makes

an evolved picture impressive is inherent in the introduced heuristics. In this way, the judgment of individual products of an artificial life simulation can be justified without appealing to subjective description. The main conclusion is that impressiveness illuminates a quantifiable facet of creative systems perhaps independent of open-endedness, one that may more deeply connect with what fascinates us about natural evolution.

Background

Because this paper introduces impressiveness, which is a measure related to open-ended evolution, this section reviews previous efforts to quantify open-ended evolution and prior investigations of concepts related to impressiveness. Novelty search, which is an approach to open-ended evolution applied in this paper's experiment, is also discussed.

Quantifying Open-Ended Evolution

In accordance with the general drive in science to formalize intuitions, there have been several attempts to quantify open-endedness (Bedau et al., 1998; Nehaniv, 2000; Standish, 2003). Such formalizations derive from intuitive features of open-ended systems, such as their drive towards diversity or complexity (Nehaniv, 2000; Standish, 2003), or their accumulation of adaptations (Bedau et al., 1998).

The dominant approach to quantifying open-ended evolution in artificial life systems is a particular measure of adaptation called *evolutionary activity statistics* (Bedau et al., 1998). The idea is that continual adaptation is a critical facet of open-ended evolution, and that persistence of traits in the face of selection is a proxy for measuring adaptation.

However, an interesting question is whether passing the activity statistics test is sufficient to equate an artificial system's creativity with that of natural evolution. Indeed, some systems have passed the test (Channon, 2001; Maley, 1999). Yet Maley (1999) acknowledges that his proposed systems will never create anything surprising and fall far short of intuitions about nature. Similarly, Channon and Damper (2000) note that in their system it eventually becomes difficult to describe what distinguishes new adaptations. In other words, passing the activity statistics test may establish open-endedness but it does not unambiguously demonstrate that a system continues to create interesting or impressive artifacts. Thus to facilitate investigating both the impressiveness of individual evolved artifacts and the tendency of artificial life simulations to create increasing impressiveness, this paper formalizes and suggests heuristics for impressiveness.

Impressiveness and Interestingness

The concept of impressiveness described in this paper also relates to the concepts of interestingness and beauty; intuitively, interesting or beautiful artifacts often tend to be impressive as well. Because they are general and important concepts, beauty and interestingness have previously been explored in diverse contexts including philosophy (Neill and Ridley, 1995), reinforcement learning (Schmidhuber, 2009), and even data mining (Geng and Hamilton, 2006).

Though they overlap in some ways, a key difference between interestingness and impressiveness is that interestingness is often tied to time-dependence or novelty (Geng and Hamilton, 2006). That is, an object that is initially found interesting may become less interesting over time due to habituation. In contrast, the formalization of impressiveness in this paper is not relative to what has been observed before. For example, the human brain will always be an impressive artifact, although by some definitions of interestingness it becomes increasingly less interesting after repeated exposure. While the term *interesting* may also sometimes be applied in a time-independent context, the term *impressiveness* explicitly disambiguates the two usages and alleviates any confusion from overlapping colloquial usage. The important point is that because the notion of impressiveness expressed here is not a relative measure it can objectively compare results *between* experiments and not only *within* them.

In addition to relating to interestingness, impressiveness might also be seen as relating in some way to beauty; for example, Schmidhuber (2009) suggests both beauty and interestingness are rooted in compressibility. The idea is that the most compressible version of an artifact may be the most beautiful. In contrast, this paper relates the concept of impressiveness to the asymmetry between ease of recognition and difficulty in creating artifacts. Importantly, it is possible that what is most impressive or beautiful about an artifact may be mostly orthogonal to compressing it; for example, aesthetic qualities such as soft, vibrant, or ornate may summarize important facets of what is appreciated about a painting without reflecting how to reconstitute it from such properties. That is, compression is typically reversible to some degree while impressive properties may be approximately one-way transformations: easy to observe but hard to create.

The next section reviews novelty search, an algorithm designed for open-ended exploration that is applied to evolving pictures in the experiment in this paper.

Novelty Search

In contrast to most EAs, which tend to converge, novelty search is a *divergent* evolutionary technique. It is inspired by natural evolution's drive to novelty, and directly rewards novel behavior *instead* of progress towards a fixed objective (Lehman and Stanley, 2008, 2011a). Thus it matches well with artificial life domains that are not motivated by a defined set of objectives. This paper will ask whether the products of novelty search are impressive.

Tracking novelty requires little change to any evolutionary algorithm aside from replacing the fitness function with a *novelty metric*, which measures how different an individual is from other individuals, thereby creating a constant pressure to do something new. The key idea is that instead of rewarding performance on an objective, novelty search re-

wards diverging from prior behaviors. Therefore, novelty needs to be *measured*.

The novelty metric characterizes how far away the new individual is from the rest of the population and its predecessors in *behavior space*, i.e. the space of unique behaviors. A good metric should thus compute the *sparseness* at any point in the behavior space. Areas with denser clusters of visited points are less novel and therefore rewarded less.

A simple measure of sparseness at a point is the average distance to the k-nearest neighbors of that point. Intuitively, if the average distance to a given point's nearest neighbors is large then it is in a sparse area; it is in a dense region if the average distance is small. The sparseness ρ at point x is given by

$$\rho(x) = \frac{1}{k}\sum_{i=0}^{k} \text{dist}(x, \mu_i), \quad (1)$$

where μ_i is the ith-nearest neighbor of x with respect to the distance metric *dist*, which is a domain-dependent measure of behavioral difference between two individuals in the search space. Candidates from more sparse regions of the behavior space then receive higher novelty scores.

If novelty is sufficiently high at the location of a new individual, i.e. above some minimal threshold ρ_{min} , then the individual is entered into the permanent archive that characterizes the distribution of prior solutions in behavior space. The current generation plus the archive give a comprehensive sample of where the search has been and where it currently is; that way, by attempting to maximize the novelty metric, the gradient of search is simply towards what is *new*, with no other explicit objective.

Once objective-based fitness is replaced with novelty, the underlying evolutionary algorithm operates as normal, selecting the most novel individuals to reproduce. Over generations, the population spreads out across the space of possible behaviors.

Instead of rewarding novel agent behaviors as in prior novelty search experiments, in this paper novelty search explores a space of *image properties*, which can be conceived as the behaviors of neural networks asked to draw pictures. In effect this approach rewards novel pictures that exhibit characteristics different from those previously encountered.

Defining Impressiveness

It is often said that the artifacts evolved by natural evolution are impressive, as are many human innovations (Darwin, 1859; Kelly, 2010). In fact, such impressiveness may be intimately connected to our appreciation of such open-ended systems. However, it is sometimes unclear whether the products of artificial systems are similarly impressive. For example, some systems have passed the evolutionary activity statistics tests designed to validate open-ended evolution (Channon, 2001; Maley, 1999) yet few researchers have accordingly concluded that recreating the dynamics of natural evolution is a solved problem. Such a discrepancy suggests that while activity statistics can successfully detect adaptation and perhaps an aspect of open-endedness, the mystery of prolific creative systems may run deeper than adaptation or open-endedness alone. In particular, an impressive open-ended system should also produce impressive artifacts. Thus a measure of impressiveness may serve as a new tool to help investigate open-ended systems.

Importantly, creating such a measure requires a definition that captures intuitions about what impressiveness means. The insight in this paper is that impressive artifacts exhibit significant design effort and that it is *easy* to recognize how difficult they were to create. To illustrate this idea, consider a gymnast performing a backflip in front of an observer.

Most observers would conclude the backflip was impressive because it takes significant strength and dexterity to defy gravity while completing a full airborne rotation and still landing squarely without falling. The general mechanisms underlying such judgments can be separated into two interrelated issues, first of mapping an observed event or artifact into an abstract description and then of judging how impressive that abstract description is. For example, the observer first recognizes the action of the gymnast as a backflip, and then evaluates how impressive a backflip is.

More specifically, the backflip is first recognized by the observer's visual system. Importantly, all that matters in observing that a backflip has occurred is that the gymnast jumps and completes a full rotation backwards in the air before successfully landing. In other words, the observer has extracted from a complex stream of sensory information a concise description that may be potentially impressive.

Once recognized, the complementary task is to judge the difficulty of this abstract description of a backflip. That is, an observer's internal understanding of physics and the athletic capabilities of most humans allows them to conclude reasonably that performing a backflip is challenging.

These two aspects combine to allow the observer to *recognize* how much effort is required to perform the action. Notice the fundamental asymmetry between recognizing and performing: It is much easier to appreciate a beautiful novel or a masterpiece than it is to create one. Interestingly, impressiveness is not a relative measure in principle. Even though it now requires less effort to create a machine that flies than it did in antiquity, the cumulative string of ideas that led to understanding flight will always be part of the true calculation of mechanized flight's impressiveness. However, in practice impressiveness may only be tractable when considered relative to a particular context (e.g. flight is not as impressive as it once was given an understanding of modern physics) or to a particular heuristic used to estimate it (e.g. re-creation effort, which is introduced later); similar practical limitations exist for other measures (Bedau et al., 1998).

Importantly, as it relates to artificial life, an impressive evolved artifact or organism will have recognizable proper-

ties that are difficult to recreate from scratch. For example, the functionality of a virtual creature might be impressive; it might locomote bipedally at a high speed, which would take many generations of evolution to achieve again. Notably, verifying an organism's speed is much simpler than creating an organism that travels at a high speed. In this way the concept of impressiveness relates to that of NP-completeness: Verifying solutions to NP-complete problems requires only polynomial computation while most researchers assume computing the solutions is impossible in polynomial time (Gasarch, 2002). Thus impressiveness can be defined as the difficulty of recreating an easily-recognized property of an artifact.

Measuring Impressiveness

The approach to investigating open-ended evolution in this paper is to measure the impressiveness of evolved artifacts. Thus this section introduces two heuristics derived from the definition of impressiveness proposed in the prior section. While it may be intractable in general to measure exactly how difficult a given property is to recreate, there are intuitive heuristics that may often reflect difficulty in practice.

The first simple such heuristic is *rarity*. That is, a property that can only be found in very small pockets of a large space may also be difficult to achieve. For example, few people are able to do backflips, which suggests it may be impressive. Similarly, few paintings are masterpieces and few novels are timeless. However, this heuristic is not without flaws because not all rare properties are hard to achieve. For instance, a person may have an odd quirk that no one else cares to acquire; though it is rare, acquiring that quirk may prove easy if attempted. Thus it is not really impressive. A more concrete example of this phenomenon can be given in the context of evolutionary algorithms. Imagine the space of all 100-digit binary numbers. Although the number consisting of all 1's is rare (occurring only once in 2^{100} possibilities), optimizing for such a property with a standard genetic algorithm is relatively trivial (Reeves, 2000). The fitness function of 1's in a given bit-string is not deceptive and is easily maximized.

Interestingly, this idea of optimizing for a particular property suggests a second, more rigorous heuristic: *re-creation effort*. If a property can be measured on a continuum, then the impressiveness of a particular level of that property can be estimated by applying a benchmark optimization algorithm to re-create that level. In other words, the difficulty for the benchmark optimizer to re-create an observed property of an evolved artifact is another way of estimating its impressiveness. Of course, the benchmark algorithm that defines the level of effort must be chosen carefully to obtain a reasonable *estimate* of the effort needed to discover a particular artifact. For example, evolving a virtual creature to reach a particular speed through a reasonable optimization algorithm may require on average a significant amount of evaluations; therefore such quick locomotion may be impressive. Relating this heuristic to the backflip example, the amount of training required for the average person to learn how to do a backflip is significant.

Both of these heuristics are applied to investigate the products of the open-ended picture evolution system that is described in the next section.

Picture Evolution Experiment

An appropriate test domain for measuring impressiveness should potentiate both open-ended discovery and achieving impressiveness. Furthermore, there can be ambiguity within the results as to whether anything of interest has really occurred. In this way, the test domain may reflect a typical artificial life system wherein interpreting its products often appeals to subjective description. The motivation is that impressiveness can instead ground such results objectively through revealing *why* particular products are impressive.

A simple such domain is evolving pictures. The phenotype space of possible pictures is vast: A square image induces c^{n^2} possibilities, where c is the number of shade gradations for a single pixel and n is the size in pixels of one dimension. Also, humans intuitively appreciate many different properties of such pictures, e.g. their dominant color, level of symmetry, or smoothness. Furthermore, some combinations of such properties may be difficult to craft, especially when they conflict. For example, a picture with a low level of smoothness that still maximizes symmetry may require some aesthetic and technical skill to draw and thus may be more impressive than other pictures.

However, because aesthetic preferences for pictures are subjective and largely variable, judging the success of a given picture evolution system may be particularly contentious. That is, people may prefer different properties of pictures, which may cause them to disagree over whether a picture-evolving system has been successful or produced anything meaningful. However, a measure of impressiveness may be able to ground statements made about evolved pictures by indicating the degree of impressiveness and *what* about particular pictures is impressive.

Following the definition of impressiveness, to fit the measures of impressiveness to picture evolution it is necessary to identify potentially impressive properties of pictures that are easily recognizable. While humans are naturally able to recognize a wide range of picture attributes, such as symmetries, similarity to real-world objects, and various aesthetic qualities, a smaller set of properties is chosen for this experiment. The motivation is to create a reasonably-sized abstract space of picture characteristics that would serve both as a basis for recognizing impressiveness and as a behavior space for novelty search to explore.

Note that although the term *space* most frequently refers to the *genotype space*, such a set of image properties is *not* the genotype space. Such image properties are *measures*

of images that will be used to help measure their impressiveness, and do not specify particular images themselves. For this experiment, eight features are chosen to capture the space of image properties, motivated by their simplicity and alignment with human recognition:

Brightness. An average of all pixel values in the picture yields a measure of a picture's brightness.

BZip2 compression. The compressibility of the image by the BZip2 algorithm gives an estimate of the picture's visual complexity.

Wavelet compression. This measure describes how compressible the image is after a wavelet transformation by counting how many coefficients are necessary to explain 95% of the image's brightness. Wavelet compression offers an alternate perspective to BZip2 on complexity.

Color variety. The standard deviation statistic is calculated over of all pixel values in a picture, giving a measure of how widely pixel values are distributed.

X-axis symmetry. This simple measure of symmetry is calculated by taking the average pixel similarity between pixels reflected over the X-axis.

Y-axis symmetry. The same measure as above is instead applied to the Y-axis.

Choppiness. The discontinuity of local neighborhoods of pixels is estimated by this measure. It is calculated as the average standard deviation of pixels over all 5x5 windows within the picture.

While the idea of impressiveness does not depend on this particular choice of picture properties, the general motivation is that such a set can facilitate aligning impressiveness with pictures visually appreciated by humans. Furthermore, they enable the evolution of impressive pictures because the trade-offs between various properties are difficult to achieve. For example, maximizing one compression measure while minimizing the other requires exploiting the differences between the underlying compression algorithms.

However, an interesting question is how to evolve pictures with such impressive properties. To do so a means of representing and evolving pictures is necessary. While there are many different representations for pictures, a well-validated method is to apply the NeuroEvolution of Augmenting Topologies (NEAT; Stanley and Miikkulainen 2002, 2004) algorithm to pictures represented by compositional pattern producing networks (CPPNs; Stanley 2007), as in Picbreeder (Secretan et al., 2011; Stanley, 2007). While the NEAT method was originally developed to evolve artificial neural networks (ANNs) to solve difficult control tasks (Stanley and Miikkulainen, 2002, 2004), it is easily adapted to evolving CPPNs because they are similar in structure to ANNs. Also, NEAT is well-suited to evolving impressive pictures because it can *complexify* CPPN topology into diverse species over generations, leading to increasingly sophisticated pictures.

In effect, CPPNs are neural networks extended to contain a variety of specially-chosen activation functions. The CPPNs in this paper take x, y coordinates as input and output the pixel brightness at that location. They facilitate images with regularities through activation functions with regular properties. For example, a Gaussian activation function by virtue of its symmetry can induce symmetric pictures and a sine function can induce pictures with elements of repetition. In this way, evolving CPPNs with NEAT can result in increasingly sophisticated images with appreciable regularities (as seen in Picbreeder; Secretan et al. 2011), which aligns well with the motivation for the experiment in this paper. Importantly, all of the experimental setups that follow apply NEAT with the same settings to evolve CPPNs; only the *reward scheme* is varied between them.

Varying the reward scheme in this way facilitates exploring the question of what type of evolutionary reward scheme is appropriate to guide this kind of open-ended search. Most approaches in EC apply objective-driven fitness functions. Yet in the huge space of potential pictures there are no inherent notions of better or worse, which usually underlies the traditional fitness-based search paradigm.

Thus with open-ended evolution in mind, a promising approach is to reward *exploring* the space of pictures through novelty search. That is, a picture is rewarded proportionally to how novel it is, i.e. how different it is from previously encountered pictures with respect to the eight picture properties (which are each scaled between 0 and 1 so that they are equally weighted). The idea is that over time as the easiest to reach points in this space are exhausted, evolution will be driven into interesting trade-offs and areas of the space that are increasingly difficult to reach. That is, novelty search may be driven to find impressive pictures. However, this sort of search has no ultimate objective other than to continually uncover new varieties of pictures and thus aligns well with the idea of open-ended evolution.

Two alternate reward schemes are also considered for comparison. First, a random search is implemented in which pictures are rewarded random fitness. The idea is to explore whether a random search, which is also open-ended in some sense because it does not attempt to prune out any possibilities from search, can also discover impressive artifacts through drift combined with NEAT's drive to complexify over time. Second, a fitness-based search is considered in which the explicit objective for each run is to re-evolve one of the rarest pictures discovered by novelty search. That is, the fitness function is to minimize distance among the salient properties (explained in the next section) from an evolved picture to the target picture. The hypothesis is that impressive artifacts may also be *deceptive* as targets and thus hard

to reach directly. If this hypothesis is true then an objective-based search to recreate such rarity may often fail to discover pictures as impressive as the target.

In this way, one aim of the experiment is to discover whether the proposed measures can make meaningful distinctions between variations in reward scheme that would naturally be expected to impact the dynamics of impressiveness. The measure's ability to make such distinctions may predict its applicability to other artificial life experiments.

Experimental Parameters

For each reward scheme 40 independent runs were conducted that ran for 500 generations each with a population size of 250. Evolved pictures were 64x64 pixels. Unlike in Picbreeder, colors in the pictures were limited to grayscale for simplicity. The dynamic threshold for adding pictures to the novelty archive was initialized to 0.5. The weight mutation power was 1.0, the chance for adding a new node was 0.05, and the chance for adding a connection was 0.1.

Results

To analyze the products of the picture evolution system, the two heuristics of rarity and re-creation effort were fitted to the domain and applied, which the next section discusses.

Recognizer Based on Rarity

To estimate how rare combinations of various values of the eight measured image properties were, ten million random CPPNs of various complexities were sampled and their properties measured. Histograms were constructed (with bins with width 0.05) for each combination of properties to estimate their joint probabilities (e.g. one such histogram would bin based on three dimensions: levels of x symmetry, wavelet compressibility, and brightness). In this way, the rarity within the space of random CPPNs of certain combinations of properties can be approximated.

To model recognition of an image's most salient features, a recognition algorithm was created that when applied to a picture would return the rarest combination of properties (e.g. the most improbable combination of properties for a particular picture might be a x-symmetry of level 0.3 and BZip2 compressibility of 0.6). In other words, the recognizer returns a summary of what is most unique about a picture, and how rare such an abstract description is (i.e. how often it occurs among randomly sampled CPPNs). Formally, the rarity of an evolved artifact is defined as $-\log(p(\alpha))$, where α is the set of salient features and $p(x)$ is a function that estimates the probability of such features occurring by chance, i.e. the probability returned by the recognizer.

In particular, the recognizer is a greedy algorithm that iterates over each combination of k features searching for the most improbable among possible combinations, starting with $k = 1$ and increasing incrementally. Because joint probabilities can only decrease when adding additional features, a control is added to ensure that adding a new feature increases rarity by at least 10 times the a priori assumption of a uniform distribution; otherwise the algorithm would terminate and return the most rare combination found so far. In this way, only the most unique properties would be considered that significantly contribute to rarity, i.e. this constraint acts as a filter to ensure concise descriptions of artifacts.

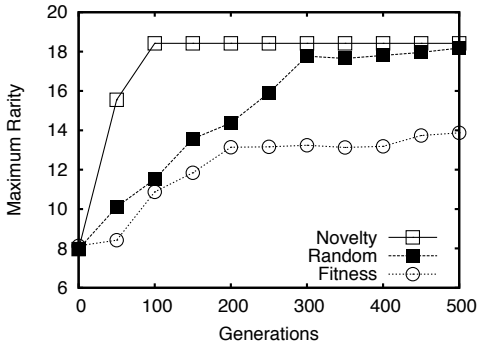

Figure 1: **Rarity of evolved images.** The maximum rarity (i.e. how infrequently similar pictures occur) of pictures from novelty search, random search, and fitness-based search is shown over 500 generations of evolution averaged over 40 independent runs. Note that a combination of image properties not present in any of the sampled CPPNs will receive a rarity of 18.4 ($2^{18.4} \approx 10,000,000$), the line to which novelty search quickly converges.

After the histograms are computed from the random CPPN samples, the recognizer algorithm is computationally inexpensive and can thus be applied to all evolved artifacts from each run at 50 generation intervals. Figure 1 shows the results of averaging the most rare picture discovered by a particular run over generations as measured by the recognizer. The main result is that rarity is able to distinguish between the different reward schemes. Novelty search is most driven towards rarity while random search more slowly discovers rarer artifacts over time (the difference is significant from generation 50 until generation 250; Student's t-test; $p < 0.001$). Novelty search also discovers significantly more rare artifacts than fitness-based search from generation 50 onwards (Student's t-test; $p < 0.001$). Interestingly, directly searching to recreate rare pictures with fitness-based search often fails due to deception.

A selection of such rare pictures found by novelty search is shown in figure 2. To aid in interpretation the combination of properties that justifies each image's rarity is returned by the observer. That is, it is possible to provide objective evidence for what is impressive about these images, instead of relying on subjective assessment as is often necessary when describing the results of an artificial life simulation. For example, picture 2a is highly compressible by BZip2 yet relatively incompressible by the wavelet algorithm, and has a low average pixel value. The result is impressive because

(a) (b) (c) (d) (e)

Figure 2: **Selection of rare pictures.** Each of these pictures discovered by novelty search was evaluated as rare by the observer because it has a combination of properties that rarely co-occur within the space of pictures.

these settings mutually conflict; generally an image is either compressible or not compressible, and incompressibility is more easily attained through wildly fluctuating pixel value (which would yield a higher average). It is also possible to learn about an encoding through observing rare artifacts: Figure 2c is rare because it is highly asymmetric along both the x and y axes and such rigidly rectangular asymmetry is not a natural bias of CPPNs (nor is it of DNA in nature).

To investigate the results of the picture evolution experiment further, the next section describes applying a more rigorous heuristic of impressiveness.

Re-creation Effort

While rarity provides one heuristic for the impressiveness of an artifact, not all that is rare is difficult to achieve. Thus conceivably the rare artifacts discovered by novelty search may require little effort to recreate, which would undermine their impressiveness.

Therefore, as a more rigorous heuristic of impressiveness, the effort required to recreate artifacts was estimated. The basic idea is to measure how much effort on average it takes to recreate a similar artifact from scratch. First, because it is computationally expensive to calculate, only the most rare picture was sampled across all 40 runs of all methods at 100 generation intervals. For each sampled picture, the observer described in the previous section derived the most rare combination of properties. Next, for each set of such observed properties five independent runs of NEAT were instantiated with those properties as an explicit objective (i.e. the fitness function was to minimize distance between the most unique properties of the target image and a candidate solution image). Each run terminated if unsuccessful after $50,000$ evaluations, or if the image *properties* were successfully recreated. The average number of evaluations required to evolve an image that would fall into the same histogram bin (i.e. allowing for error of 0.05 in any given property) was then recorded as an estimate of the effort required to recreate a similar picture.

Figure 3 shows these results for all three variants, which reinforce the results from measuring rarity in the previous section by distinguishing the reward schemes in the same order. In particular, novelty search is distinguished from random search for generations 100 and 200, and from fitness-based search for all generations after zero (Student's t-test; $p < 0.001$). It is interesting that random search demonstrates a drive towards impressiveness (which may result from NEAT's complexification mechanism), although novelty search most quickly evolves artifacts that exceed the upper extreme of the test's range ($50,000$ evaluations).

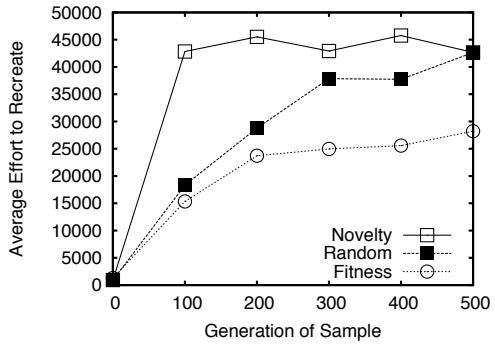

Figure 3: **Effort to recreate image properties.** The average effort (i.e. the number of evaluations) necessary on average to recreate the rarest images (with fitness-based search) sampled at 100 generation intervals from each run of novelty search, random search, and fitness-based search is shown. Note that the measure has a ceiling of $50,000$ evaluations, which may mask continuing growth of re-creation effort for both novelty search and random search.

Additionally, a significant correlation (0.673) was measured between paired samples of rarity and re-creation effort ($p < 0.0001$; Kendall's tau coefficient), indicating that the two heuristics are strongly related, which supports their derivation from the same definition.

Discussion

From a practical perspective the definition and heuristics of impressiveness introduced in this paper facilitate making distinctions among variations of evolutionary systems and providing objective statements about their products. Novelty search, which is designed explicitly to achieve open-ended exploration, climbs the ladder of impressiveness most steeply, as would be intuitively expected. However, on a deeper level impressiveness yields an alternate perspective on the goals of open-ended evolution.

That is, perhaps meeting the challenge of unbounded open-endedness, which is often assumed to correlate with intuitions about natural evolution's greatness, is instead a necessary but not sufficient condition to yield increasingly impressive products. In other words, increasing impressiveness may be a more inherently meaningful goal than open-endedness alone insofar as it more deeply abstracts what we appreciate about natural evolution: its impressive products.

Furthermore, the results in this paper and prior work with non-objective search processes (such as novelty search and Picbreeder) suggest that objective-based search is deceived by increasingly ambitious or impressive objectives (Lehman and Stanley, 2011a,b; Woolley and Stanley, 2011). Thus, an interesting possibility is that open-endedness may be important to evolving increasingly impressive artifacts solely to circumvent deception. That is, seeking impressiveness convergently may be fruitless because of the inherent difficulty in predicting a priori what paths through any search space will lead to great achievement. Such a possibility hints at a potential deeper understanding of open-ended creativity.

Future work will investigate the hypothesis that systems previously passing the evolutionary activity statistics tests will not exhibit unbounded impressiveness, highlighting where the two measures may differ.

Conclusion

Motivated by the possible gap between open-endedness and impressiveness in some artificial life simulations, this paper introduced the idea of quantifying the impressiveness of evolved artifacts. Heuristic measures of impressiveness derived from a novel definition were applied to an open-ended picture evolution system to characterize the effect of different reward schemes on impressiveness and to examine individual evolved products. The conclusion is that impressiveness is a new tool for investigating the products of open-ended systems that presents an alternate perspective on the goals of open-ended evolution.

Acknowledgements

This research was supported by DARPA and ARO through DARPA grant N11AP20003 (Computer Science Study Group Phase 3) and US Army Research Office grant Award No. W911NF-11-1-0489. This paper does not necessarily reflect the position or policy of the government, and no official endorsement should be inferred.

References

Bedau, M., Snyder, E., Brown, C. T., and Packard, N. H. (1997). A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In Husbands, P. and Harvey, I., editors, *Proceedings Of The Fourth European Conference on Artificial Life*, pages 125–134. MIT Press.

Bedau, M. A., Snyder, E., and Packard, N. H. (1998). A classification of longterm evolutionary dynamics. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Proceedings of Artificial Life VI*, pages 228–237, Cambridge, MA. MIT Press.

Channon, A. (2001). Passing the alife test: Activity statistics classify evolution in geb as unbounded. In *Proceedings of the European Conference on Artificial Life(ECAL-2001)*. Springer.

Channon, A. D. and Damper, R. I. (2000). Towards the evolutionary emergence of increasingly complex advantageous behaviours. *International Journal of Systems Science*, 31(7):843–860.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life*. Murray, London.

Gasarch, W. (2002). The P =? NP poll. *Sigact News*, 33(2):34–47.

Geng, L. and Hamilton, H. (2006). Interestingness measures for data mining: A survey. *ACM Computing Surveys (CSUR)*, 38(3):9.

Kelly, K. (2010). *What technology wants*. Viking Press.

Lehman, J. and Stanley, K. O. (2008). Exploiting open-endedness to solve problems through the search for novelty. In Bullock, S., Noble, J., Watson, R., and Bedau, M., editors, *Proceedings of the Eleventh International Conference on Artificial Life (ALIFE XI)*, Cambridge, MA. MIT Press.

Lehman, J. and Stanley, K. O. (2011a). Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation*.

Lehman, J. and Stanley, K. O. (2011b). Novelty seach and the problem with objectives. In *Genetic Programming in Theory and Practice IX (GPTP 2011)*, chapter 3, pages 37–56. Springer.

Maley, C. C. (1999). Four steps toward open-ended evolution. In *Proceedings of the Genetic and Evolutionary Computation Conference(GECCO-1999)*, volume 2, pages 1336–1343, Orlando, Florida, USA. IEEE Press.

Nehaniv, C. (2000). Measuring evolvability as the rate of complexity increase. In Maley, C. and Boudreau, E., editors, *Artificial Life VII Workshop Proceedings*, pages 55–57.

Neill, A. and Ridley, A. (1995). *The philosophy of art: readings ancient and modern*, pages 98–239. McGraw-Hill.

Reeves, C. (2000). Fitness landscapes and evolutionary algorithms. In *Artificial Evolution*, pages 3–20. Springer.

Schmidhuber, J. (2009). Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Anticipatory Behavior in Adaptive Learning Systems*, pages 48–76.

Secretan, J., Beato, N., D'Ambrosio, D., Rodriguez, A., Campbell, A., Folsom-Kovarik, J., and Stanley, K. (2011). Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary Computation*, 19(3):373–403.

Standish, R. (2003). Open-ended artificial evolution. *International Journal of Computational Intelligence and Applications*, 3(167).

Stanley, K. (2007). Compositional pattern producing networks: A novel abstraction of development. *Genetic Programming and Evolvable Machines*, 8(2):131–162.

Stanley, K. O. and Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10:99–127.

Stanley, K. O. and Miikkulainen, R. (2004). Competitive coevolution through evolutionary complexification. 21:63–100.

Taylor, T. and Hallam, J. (1998). Replaying the tape: An investigation into the role of contingency in evolution. In Taylor, C., Langton, C., and Kitano, H., editors, *Proceedings of Artificial Life VI*, pages 256–265.

Woolley, B. G. and Stanley, K. O. (2011). On the deleterious effects of a priori objectives on evolution and representation. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2011)*. ACM.