

On the Deleterious Effects of A Priori Objectives on Evolution and Representation

In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2011)*. New York, NY: ACM

Brian G. Woolley
Department of EECS
University of Central Florida
Orlando, FL 32816, USA
brian.woolley@ieee.org

Kenneth O. Stanley
Department of EECS
University of Central Florida
Orlando, FL 32816, USA
kstanley@eecs.ucf.edu

ABSTRACT

Evolutionary algorithms are often evaluated by measuring and comparing their ability to consistently reach objectives chosen *a priori* by researchers. Yet recent results from experiments without explicit *a priori* objectives, such as in Picbreeder and with the novelty search algorithm, raise the question of whether the very act of setting an objective is exacting a subtle price. Nature provides another hint that the reigning objective-based paradigm may be obfuscating evolutionary computation's true potential; after all, many of the greatest discoveries of natural evolution, such as flight and human-level intelligence, were *not* set as *a priori* objectives at the beginning of the search. The dangerous question is whether such triumphs only result *because* they were not objectives. To examine this question, this paper takes the unusual experimental approach of attempting to re-evolve images that were *already* once evolved on Picbreeder. In effect, images that were originally discovered serendipitously become *a priori* objectives for a new experiment with the *same algorithm*. Therefore, the resulting failure to reproduce the very same results cannot be blamed on the evolutionary algorithm, setting the stage for a contemplation of the price we pay for evaluating our algorithms only for their ability to achieve preconceived objectives.

Categories and Subject Descriptors:

I.2.6 [Artificial Intelligence]: Learning

General Terms:

Keywords: Fitness, deception, non-objective search, stepping stones, representations, indirect encoding

1. INTRODUCTION

Evolutionary algorithms (EAs) are often tested on benchmarks to assess their ability to reach a particular objective in the search space. Popular types of benchmarks include optimization [2], function approximation [27], and control (such as in neuroevolution [6]). Reinforcing the field's focus on objective-targeted optimization, theoretical analyses often focus on the probability that EAs will converge to the objective [3, 4, 17]. A fundamental assumption behind this objective-focused paradigm in evolutionary computa-

tion (EC) is that the value of EAs lies in their ability to reach objectives that we set for them *a priori*. Paired with this assumption is the idea that when an evolutionary algorithm *consistently fails* to reach its intended objective, then it is not suited for the task. Yet what if this central assumption is wrong? What if consistently reaching the objective could actually obfuscate a deeper underlying pathology, while consistently failing to reach the objective ultimately bears little on the ability of the EA to produce impressive results in general?

While perhaps paradoxical, these questions highlight a delicate uncertainty in EC about its relationship to natural evolution, where in many remarkable phenotypes were discovered even though none of them were explicitly expressed as *a priori* objectives for the process. Is it possible that we are judging our algorithms wrongly?

To gain a fresh perspective on this question this paper takes an unusual experimental approach: The objective of the EA is to *re-evolve* images that were *already* evolved in Picbreeder [21, 22], an online service on which users collaboratively evolve images through interactive evolution [26]. Such an experimental approach yields a unique perspective because many interesting images have already been evolved on Picbreeder by its users; while these images were not specified *a priori* as objectives when Picbreeder was first introduced, in this paper they *become* objectives for the very same algorithm and representation already inside Picbreeder. That way, even if the EA fails to reproduce its own original results, the conclusion that the algorithm or representation is insufficient to produce the objective is *precluded*, because the algorithm *did* already produce the objective, only it was originally discovered serendipitously before it was chosen as a benchmark objective. Through this experiment, we will discover that just the act of setting an objective triggers a chain of unintended consequences that confound the usual conclusions drawn from such benchmarks.

In particular, the main result disclosed in this paper is that the algorithm inside Picbreeder, NeuroEvolution of Augmenting Topologies (NEAT [24,25]) and the representation inside Picbreeder, compositional pattern-producing networks (CPPNs [23]), consistently fail to reproduce the results that they already produced. For the simpler images attempted, the failure is in the quality of the solution, yielding much larger representations than necessary; for the more complex images, the failure is in obtaining the solution at all.

However, the most revealing focus of the analysis is on *why* these failures occur so universally. Because NEAT in fact already discovered the targets, we are forced towards deeper insight into the phenomenon of failure than simply blaming NEAT, as would be customary. In fact, it turns out that it is the very practice of making the images objectives that ultimately causes them not to be rediscovered properly. Not only does evaluating fitness against an objective potentially push it in the wrong direction through deception, but

even when it still manages to right itself, the result is often a destructive effect on *representation* that has received little attention before this study.

This insight is sobering because the vast majority of empirical experiments in EC assess their results based on benchmarks with a priori objectives [2, 8, 13, 20, 24, 27] and theory often focuses on whether targets can be reached [3, 17]. Thus the major implication is that a change in thinking may be warranted about *how EAs should be judged*. Furthermore, on the positive side, the analysis hints at the kinds of situations that ultimately encourage efficient and elegant representations to evolve. As the discussion so far implies, these ideal situations are precisely when there is no specified objective (as in Picbreeder), suggesting that the way we traditionally use these algorithms may not ultimately be allowing them to exhibit their full potential.

2. BACKGROUND

This section reviews foundational work in EC that motivates and underlies the experiments in this paper.

2.1 Evolutionary Computation

In evolutionary computation (EC), it is traditional to evolve the population iteratively by evaluating its members with respect to the objective [2, 4, 7, 8]. While initial populations are often largely unfit with respect to the objective, some individuals are generally more fit than others. Inspired by the Darwinian principle of natural selection [8], the variation in the fitness of individuals provides a gradient to guide the search. In this way, EC explores multiple candidate solutions in parallel, thereby providing hope that it may avoid the trap of local optima.

Although EC can be divided into a number of branches or sub-communities, they all follow this general framework, i.e. a cycle of evaluation, selection, and mutation is applied repeatedly to shape a population with respect to an objective [2, 4]. For example, *genetic algorithm* (GA) approaches generally optimize strings of numbers that represent parameters in a problem domain [8]; *genetic programming* (GP) techniques evolve computer programs as trees of operators and operands to solve computational tasks [13]; *evolutionary programming* (EP) searches for optimal solutions by evolving the numerical parameters for fixed computational structures like neural networks and program trees [20]; and *neuroevolution* (NE) evolves the structure and connection weights for artificial neural networks to perform control and decision-making tasks [11, 24, 25].

While the stochastic elements of EAs provide the ability to escape local optima, the general approach remains an objective-based search that selects individuals based on how closely they resemble the a priori objective. A major challenge for this approach is thus to reward the intermediate steps that are required to reach the optimal solution. In this way, landscapes induced by objective (e.g. fitness) functions are often deceptive [9, 16, 18]. The problem is that the objective function does not necessarily reward the *stepping stones* in the search space that ultimately lead to the objective. In cases where the objective is too complex to express directly, *incremental evolution* is often applied to reward the evolution of the stepping stones necessary to reach the final solution [10]. The challenge of encompassing the most natural stepping stones in the objective function will prove an essential factor in the negative results reported in this paper.

2.2 Non-Objective Search

A fundamental dilemma with traditional approaches is that crafting an effective fitness function is akin to understanding the fitness landscape or knowing the stepping stones a priori [5, 28]. This

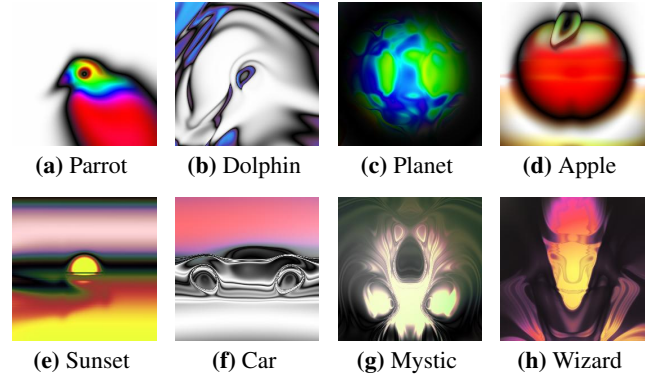


Figure 1: Images Evolved on Picbreeder. These images were interactively evolved by a community of human users with no explicit objective. They demonstrate the system’s ability to discover interesting and meaningful images.

problem even applies to coevolution, where the objective is effectively implicit [5]. In all cases, the need to identify an effective fitness gradient becomes increasingly difficult as objectives become more ambitious because the intermediate steps to the solution are less likely to be known [5]. Highlighting this problem, Lehman and Stanley [14] demonstrated that searching without regard to the objective, i.e. searching only for novel behavior, is more effective at discovering solutions in some deceptive domains than rewarding objective performance. In this approach, called *novelty search*, stepping stones are often preserved because they are novel, whereas otherwise they would have been thrown away for being far from the objective. Results with novelty search demonstrated that it finds neural networks that solve deceptive mazes more frequently, more quickly, and at lower genomic complexities than an equivalent objective-based approach [14, 15]. The compressed representations reported for novelty search inspired the investigation in this paper into how searching with and without objectives affects representation.

Yet novelty search is not the only way to implement an evolutionary search without a priori targets. Another effectively non-objective domain is Picbreeder [21, 22], which is leveraged in this paper to compare representations from objective and non-objective searches. Picbreeder is a distributed community of users that interactively evolve pictures through *interactive evolutionary computation* (IEC), which means selecting images that are appealing to produce a new generation of images [26]. The idea of IEC originated in Richard Dawkins’ book, *The Blind Watchmaker*, in which he described a simple program called *Biomorphs* that is meant to illustrate evolutionary principles [1]. In Picbreeder it is expanded to become *collaborative* such that users on Picbreeder can continue to evolve images previously evolved by other users. The collection of images generated by Picbreeder is significant because it demonstrates how a group of individuals working without a formal unified objective can discover *attractive* and *interesting* areas in the vast desert of all possible images; some such images are shown in figure 1. Additionally, the quality of such a serendipitous approach to evolution is evident in the diverse phylogeny of images that have emerged, the compactness of their representations, and the speed (i.e. low number of generations) with which meaningful images are discovered. A crucial aspect of this result for the purposes of the present study is that the system *as a whole* has no unified a priori objective or objectives. While individual users may sometimes (and sometimes not) arrive with their own objectives, the combination of all users branching off each other is not working towards

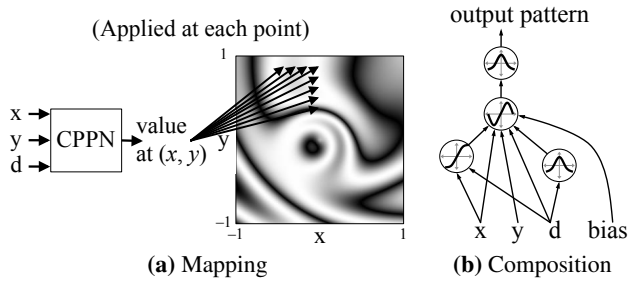


Figure 2: CPPN Encoding. (a) The CPPN takes arguments x and y , which are coordinates in a two-dimensional space. The additional input d is the distance between (x, y) and the image center, which provides a helpful bias. When all the coordinates are drawn with an intensity corresponding to the output of the CPPN, the result is a spatial pattern that can be viewed as a phenotype whose genotype is the CPPN. Internally, the CPPN (b) is a graph that determines which functions are connected. As in an ANN, the connections are weighted such that the output of a function is multiplied by the weight of its outgoing connection. The CPPN in (b) actually produces the pattern in (a).

any unified objective in particular. This fact will turn out critical to the representations that ultimately evolve.

Therefore, because the representations evolved by Picbreeder are compact and elegant, it is chosen as the vehicle for generating target images in the study in this paper. That way, it is possible to compare the newly-evolved representations with the ones in Picbreeder and see the effect of objectivizing what was initially largely serendipitous. Thus, the next section explains how Picbreeder works.

2.3 Picbreeder

The online service at picbreeder.org allows users to collaboratively evolve images. Users evolve images in Picbreeder by selecting ones that appeal to them from among a set of candidates to produce a new generation. As this process is repeated, the individual images in the population evolve to satisfy the user. Once satisfied, the user can *publish* his or her image to the Picbreeder site. Sharing such work with the community then allows others to continue evolving already-published images to form new and more intricate designs [21, 22], which is called *branching*.

Each image in Picbreeder is indirectly encoded by a variant of a neural network called a *compositional pattern-producing network* (CPPN [23]). The idea behind CPPNs is that geometric patterns can be encoded by a *composition of functions* that are chosen to represent common regularities. Given a function f and a function g , a composition is defined as $f \circ g(x) = f(g(x))$. Thus, a set of simple functions can be composed into more elaborate functions through hierarchical composition, e.g. $f \circ g(f(x) + g(x))$. For example, composing the Gaussian function, which is symmetric, with any other function results in a symmetric pattern. The internal structure of a CPPN is represented as a weighted network, similar to an artificial neural network (ANN), that denotes which functions are composed and in what order. The appeal of this encoding is that it can represent a pattern with regularities such as symmetry, repetition, and repetition with variation through a network of simple functions that can be evolved by existing methods for evolving ANNs.

Images are rendered from CPPNs in Picbreeder by querying the network as the function $\text{CPPN}(x, y, d)$ to obtain the grayscale value of the pixel located at (x, y) in the image (figure 2).¹ The extra input d is the *distance from the center* to the (x, y) location being queried, which gives the CPPN a sense of radial symmetry that

¹While some Picbreeder CPPNs also generate color images, the images chosen as targets in this study are all grayscale.

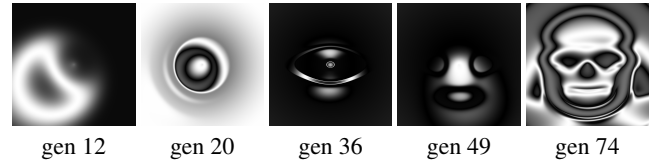


Figure 3: Stepping stones to the Skull. This image illustrates the progression of seemingly unrelated images that led to a seminal Picbreeder image (i.e. the *Skull*) in just 74 generations. As a metaphor for solving ambitious problems, such an example demonstrates the importance of preserving stepping stones even if they do not resemble an a priori objective.

provides a bias towards appealing images [21]. Because CPPNs are a composition of continuous functions in a geometric space they provide a compact representation of the image at any resolution.

Picbreeder evolves images by evolving the underlying CPPN as if it were an ANN. That way, the *NeuroEvolution of Augmenting Topologies* (NEAT) approach [24, 25] can facilitate the evolutionary step between generations. More specifically, the NEAT algorithm starts with a population of simple CPPNs and *complexifies* them over generations by adding new nodes and connections through structural mutations. By evolving networks in this way, the topology of the network does not need to be known a priori. As a result of this process, the initial population of Picbreeder images begins as simple patterns represented by networks with just five connections and one hidden node. As the underlying networks add complexity, features and structures emerge in the resulting images that could not be expressed by the simpler CPPNs. When interesting or meaningful changes occur, the user selects the images that will reproduce to form the next generation. In this way, users can rapidly move from simple patterns towards a higher-dimensional space by rewarding regularities and structures that have meaning to them. At the same time, the variable size of the genome means that some evolved representations are better than others, i.e. by taking less space. The kind of evolutionary process that yields the most compact representations is a focus of this paper.

A further important aspect of the experiments in this paper is that they are *also* implemented with NEAT (in fact, the same software as in Picbreeder is used). That way, the same algorithm and representation that produced the original targets will later attempt to produce them again. For a complete overview of NEAT see Stanley and Miikkulainen [24, 25].

3. MOTIVATION

While the idea of searching without an objective may at first seem unusual, even a brief experience with Picbreeder shows that a serendipitous excursion through interesting parts of the search space does not require any specific objective. On the way, stepping stones and key innovations that appear interesting in their own right are discovered that can be elaborated later (perhaps by another user) to reach ambitious ends, even when the intermediate steps do not resemble the ultimate destination.

Furthermore, practical experience with Picbreeder has shown the futility of setting out to evolve a particular type of image, e.g. a flower or a butterfly [21, 22]. In contrast, novice users that are encouraged to evolve patterns without an a priori expectation frequently find appealing images within 10 to 30 generations (often after branching from images evolved by other users that do not resemble the newer result). It turns out that evolving images in this way works because it tends to reward structures that become stepping stones to other meaningful images even if the stepping stones do not resemble their descendants. An example of this phenomenon is the lineage of images in figure 3 that ultimately led

to the *Skull*, a seminal Picbreeder image. Its predecessors, which resemble e.g. a crescent, a donut, and a dish, do not hint at the significant discovery to come, yet were nevertheless essential to reaching it. Users involved in this lineage thus selected these images *for their own appeal* rather than because they were searching explicitly for a skull. As a metaphor for finding solutions to ambitious problems, the value of such important intermediate steps often cannot be known in objective-driven search when they are first discovered because essential innovations do not necessarily resemble the given objective.

The emergence of novel forms from common ancestors can be attributed to the ability of evolution to elaborate on existing regularities and to adapt (i.e. through exaptation) the structures developed in existing images. The evolution of an image with meaningful semantic structure (e.g. the features of a face) is significant because a variety of conflicting personal preferences are directing the selection process in Picbreeder. An evolutionary approach that selects individuals for their interesting qualities allows early decisions about what is important, e.g. bilateral symmetry, to become the founding principles for subsequent generations [23]. Individuals with interesting qualities may then reveal themselves as stepping stones to more intricate discoveries that are radically different than the original, yet still elaborate on an underlying principle.

The hypothesis of this paper is that an objective-based search will construct a piecewise solution that fails to embody key regularities in the problem domain because the fitness function only rewards incremental improvements that *resemble* the objective, which is highly restrictive. By favoring short-term gains rather than developing a broad variety of innovations, early decisions about how to construct the underlying representation of solutions will fail to capture the key organizational concepts in the problem domain.

The experiment described next makes these considerations concrete by showing in practice how setting an objective distorts the search and corrupts the representation, even with the very same algorithm and representation with which the targets were originally discovered. The key lesson will be that *how* something is discovered matters, and discovery through objective optimization is often the wrong way to do it.

4. EXPERIMENT

The aim of this experiment is to investigate how pursuing a singular objective impacts the underlying representation of the solution. To achieve this aim, a population of image-generating CPPNs (i.e. with inputs x , y , and d , as in Picbreeder) is evolved towards one of the six target images (which span a range of different complexities) in figure 4. It is important to note that each target image was originally evolved by a human user or chain of users on Picbreeder. Thus the task of evolving CPPNs toward these objectives should be feasible because the targets are *known* to exist in the space of solutions and were previously evolved.

To ensure that a direct comparison can be made between the representation of solutions evolved serendipitously and those evolved as objectives, the automated evolution described in this section has the same operational parameters as Picbreeder wherever possible. The key difference is that in interactive evolution, a human user selects the individuals that will become the parents of the next generation. Thus this experiment adds an evaluation and selection process to replace the human evaluator.

All experiments were run with a version of the public domain ANJI NEAT package [12] augmented to include the Picbreeder CPPN network classes and mutation code [22] (like ANJI, Picbreeder is written in Java so the code could be combined). As in Picbreeder, the available CPPN activation functions were sine,

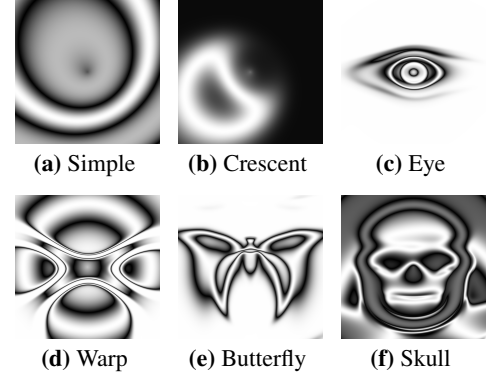


Figure 4: Target images. Each image was originally evolved by human users in Picbreeder. The difficulty of re-evolving toward these images with the same algorithm and representation will demonstrate that *how* something is discovered matters.

cosine, Gaussian, identity, and sigmoid. Recurrent connections within the CPPN were not enabled and signed activation was enforced. The network output w was in the range $[-1 \dots 1]$ and the corresponding grayscale value was calculated as $256(1 - |w|)$. To facilitate automated evolution, unlike Picbreeder, the population size was increased from 15 to 150 and each run terminated at generation 30,000. The speciation threshold (δ_i) in NEAT was 0.2 and the compatibility modifier was 0.3. These parameters were found to produce similar results with moderate variation.

To automate the selection process, the fitness of each evolved candidate image is calculated based on how closely it matches the target image. In effect, the target images become the objective and thereby serve as metaphors for objectives of different levels of complexity. To compare two images and calculate fitness, each is defined by a feature set that includes the grayscale value at each pixel location (at 128×128 pixels) and the gradient between adjacent pixel values. The candidate feature set is then scaled to correspond with the normalized target feature set. In this way, the candidate and target images can be compared by calculating the *degree of difference* [19] between corresponding features. The degree of difference, $d(c, t)$, between a particular candidate feature c and the corresponding target feature t is defined as

$$d(c, t) = 1 - e^{-\alpha|c-t|}, \quad (1)$$

where $\alpha = 5$ is a modulation parameter. From this equation, images can be described by their feature sets, wherein elements in the candidate feature set, $C = \{c_1, \dots, c_n\}$, can be compared against the target feature set, $T = \{t_1, \dots, t_n\}$. Thus the error between the candidate and the target feature sets, $\text{err}(C, T)$, is calculated as

$$\text{err}(C, T) = \frac{1}{n} \sum_{i=1}^n d(c_i, t_i), \quad (2)$$

which is the average degree of difference across feature sets C and T , where n is the total number of features. Finally, the fitness of the candidate, $f(C)$, is assigned as

$$f(C) = 1 - \text{err}(C, T)^2, \quad (3)$$

such that taking the mean-squared error (MSE) generalizes large errors and emphasizes the importance of details.

Inevitably the fitness function in this experiment will be scrutinized for its effectiveness. Some will ask whether a better such function for comparing images could have been chosen. However, it will turn out that ultimately the problem is that no image comparison, no matter how good, can really reward stepping stone images

because stepping stones to the target do not resemble the target itself. Thus the problem here is significantly deeper than simply finding a better image-matching heuristic. Nevertheless, to validate the fitness function as a reasonable search heuristic, a preliminary experiment evolves a population of CPPNs towards a randomly-generated image with five connections and one hidden node, i.e. the *Simple* target image in figure 4a. The validation experiment differs from the main experiment in that the correct solution topology for the CPPN is given (i.e. it is taken from the previously-evolved target), and only the connection weights must be evolved (unlike in normal NEAT wherein both weights and structure evolve together). Thus evolution must match the target image by tuning the five connection weights. The results of this preliminary validation of the fitness function showed that the solution threshold (explained below) was reached by 19 of 20 runs in 353 ± 687 (median = 135) generations on average. In this way, the validation experiment demonstrates that the fitness function is a reasonable method for comparing images in the population to the target image.

5. RESULTS

The key question is what happens when NEAT attempts to re-evolve as objectives in an automated search the images that were already evolved by Picbreeder users. For consistency, a run is considered *successful* if the fitness score is greater than 0.75 (out of 1.0 maximum). This threshold corresponds to an average error of 5.75% between the source feature set and the target feature set and was verified as reasonable by a qualitative review of the resulting images above this threshold.

To show the contrast between images re-evolved as objectives as opposed to discoveries made on Picbreeder, table 1 compares the objective-based results against the statistics of the target images when they were originally evolved. The names that refer to the different targets in this section are given in figure 4. Of runs evolving to the *Simple* target, 14 of 20 reached the solution threshold in $3,774 \pm 5,902$ (median = 965) generations on average (recall that the simple target is from generation 1). Of these, the objective-based solutions were significantly larger (19.5 ± 2.7 functions and 23.4 ± 4.2 connections) than the original Picbreeder discovery (6 functions and 5 connections). Similarly, for runs evolving toward the *Crescent*, 11 of 20 reached the solution threshold in $3,500 \pm 3,298$ (median = 2,320) generations (as opposed to only 12 generations for the original discovery of the *Crescent*). Of these, the representation was again significantly bloated (20.3 ± 3.2 functions and 26.4 ± 4.4 connections) compared to the target evolved in Picbreeder (7 functions and 7 connections). When evolving toward the *Eye*, only one of 20 runs reached the solution threshold. This particular run reached the solution criterion at generation 4,840 (as opposed to 12 generations to find the original in Picbreeder). Furthermore, the resulting solution is once again more complex (18 functions and 24 connections) than the target (10 functions and 16 connections).

Beyond these simpler images, at a certain level of complexity it becomes too hard to come even close to hitting the targets: All attempts to recreate the *Warp*, the *Butterfly*, and the *Skull* failed to produce comparable solutions.

A set of champions from the 20 attempts to reach each of the six targets are shown in figure 5. To satisfy space constraints and to provide a comprehensive sampling of the typical results of evolving to each target, figure 5 shows the final result from the ten odd-numbered runs for each target (out of 20 for each). The even runs (not shown) generally follow a similar pattern. The size of the CPPNs of these champions are also shown. In cases labeled as *failed*, the run could not meet the solution criteria even after 30,000

Target	Solved	Generations	Functions	Connections
Simple	14	$3,774 \pm 5,902$ (1)	19.5 ± 2.7 (6)	23.4 ± 4.2 (5)
Crescent	11	$3,500 \pm 3,298$ (12)	20.3 ± 3.2 (7)	26.4 ± 4.4 (7)
Eye	1	4,840 (12)	18 (10)	24 (16)
Warp	0	- (5)	- (9)	- (12)
Butterfly	0	- (90)	- (25)	- (75)
Skull	0	- (74)	- (23)	- (57)

Table 1: Image evolution results. The number of runs (out of 20) that reached the solution threshold, the number of generations required to reach a solution, and the CPPN solution complexity (i.e. the number of functions and connections) are shown. Values are averaged over runs that achieved the solution threshold. For comparison, the same results for the original discoveries of each target in Picbreeder are shown in parenthesis.

generations. These pictorial results give a qualitative sense of how difficult it is for NEAT to reach the targets it once evolved previously under different conditions. Based on these results, the next section discusses the destructive impact of objective-based search on representation.

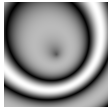

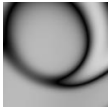

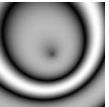

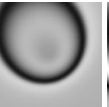
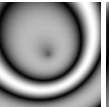

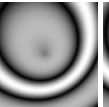
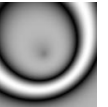
6. DISCUSSION

The results document nothing less than the systematic failure of the objective-driven search. It could not come close to reproducing most of the images. On the other hand, although it could reasonably reproduce the simplest two images (i.e. *Simple* and *Crescent*), the re-evolved solutions in those cases contain two to four times more structure than the CPPNs of the original discoveries. Taken together, these dismal results signify a serious underlying pathology; because we know that the target images were evolved in the first place, it should have been possible to reproduce them and to reproduce them efficiently.


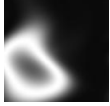


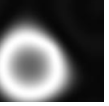
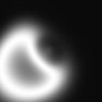
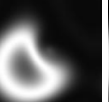
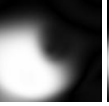
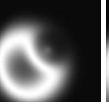
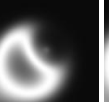

The key question is whether these results extend beyond the confines of this particular study to imply something about objective-based search in general. The aim of this section is to argue that they do. However, a natural reaction to such poor performance is to dismiss it by pinning it to specific shortcomings of the algorithm and representation or the experimental setup. Yet the unusual experimental setup, in which targets were chosen that were already evolved by the same algorithm and representation, makes it difficult to dismiss the results so easily.

The first objection to drawing general implications is that something might be wrong with NEAT or CPPNs that could be rectified in a different setup. However, it is difficult e.g. to argue that NEAT and CPPNs are somehow ill-equipped to evolve skulls when the *Skull* has only ever been evolved by NEAT and CPPNs. Not only that, but the *Skull* was originally discovered in only 74 cumulative generations with a population of 15 on Picbreeder, which makes it hard to argue coherently that it is "difficult" for NEAT to discover. Yet 30,000 generations was not sufficient to rediscover it when it was the objective. Furthermore, although there may be better algorithms, NEAT generally has a good record in a variety of domains [24, 25, 27], diminishing the possible argument that it is a kind of straw man chosen for its weaknesses.












The other potential objection is to the experimental setup. Perhaps equations 1–3 could somehow better compute image similarity. Yet this objection depends on the assumption that the basic challenge in evolving to a target is to formalize a good *similarity metric*, which turns out not to be the real problem. To see why, observe the stepping stones in figure 6 that led to the discovery of the original *Eye* in just 12 generations (recall that the only successful rediscovery of the eye out of 20 attempts took 4,840 generations). The problem is that the early stepping stones, such as in the

Simple	Run 1	Run 3	Run 5	Run 7	Run 9	Run 11	Run 13	Run 15	Run 17	Run 19
										
6f, 5c 1 gen	19f, 27c failed	18f, 25c failed	19f, 22c 3,570 gen	17f, 23c 440 gen	18f, 26c 1,200 gen	19f, 23c failed	16f, 24c 1,090 gen	20f, 23c failed	16f, 20c 840 gen	25f, 20c 9,320 gen


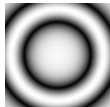




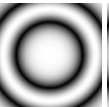
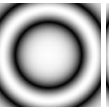
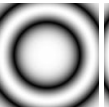
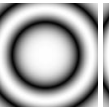

(a) Champions from the *Simple* runs.

Crescent	Run 1	Run 3	Run 5	Run 7	Run 9	Run 11	Run 13	Run 15	Run 17	Run 19
										
7f, 7c 12 gen	18f, 24c failed	19f, 23c 1,300 gen	20f, 26c failed	17f, 22c failed	20f, 27c 9,270 gen	24f, 26c 10,760 gen	20f, 33c failed	19f, 24c 3,380 gen	20f, 24c 1,580 gen	12f, 17c 1,830 gen









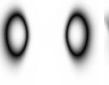


(b) Champions from the *Crescent* runs.

Eye	Run 1	Run 3	Run 5	Run 7	Run 9	Run 11	Run 13	Run 15	Run 17	Run 19
										
10f, 16c 12 gen	22f, 24c failed	18f, 18c failed	19f, 20c failed	22f, 21c failed	24f, 30c failed	18f, 24c 4,840 gen	17f, 18c failed	18f, 15c failed	16f, 21c failed	22f, 23c failed





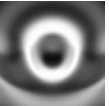
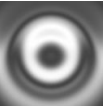
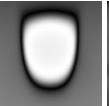
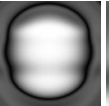
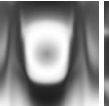
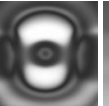

(c) Champions from the *Eye* runs.

Warp	Run 1	Run 3	Run 5	Run 7	Run 9	Run 11	Run 13	Run 15	Run 17	Run 19
										
9f, 12c 5 gen	18f, 17c failed	18f, 22c failed	24f, 23c failed	19f, 20c failed	19f, 24c failed	22f, 20c failed	20f, 21c failed	22f, 26c failed	14f, 17c failed	17f, 18c failed

(d) Champions from the *Warp* runs.

Butterfly	Run 1	Run 3	Run 5	Run 7	Run 9	Run 11	Run 13	Run 15	Run 17	Run 19
										
25f, 75c 90 gen	22f, 27c failed	21f, 27c failed	22f, 25c failed	20f, 28c failed	18f, 23c failed	21f, 27c failed	27f, 34c failed	22f, 25c failed	24f, 29c failed	20f, 28c failed

(e) Champions from the *Butterfly* runs.

Skull	Run 1	Run 3	Run 5	Run 7	Run 9	Run 11	Run 13	Run 15	Run 17	Run 19
										
23f, 57c 74 gen	20f, 24c failed	20f, 29c failed	19f, 24c failed	22f, 28c failed	21f, 28c failed	16f, 22c failed	21f, 27c failed	23f, 29c failed	18f, 25c failed	25f, 28c failed

(f) Champions from the *Skull* runs.

Figure 5: Image evolution results. Results from the ten odd-numbered runs to each target are shown (as they were evaluated, at 128×128 pixels). The even runs (not shown to satisfy space constraints) generally follow a similar pattern. The CPPN complexity (i.e. the number of functions and connections) is shown for each image along with the generation when the solution threshold was reached. Runs for which no solution was reached are labeled *failed*. This pictorial perspective demonstrates the difficulty that NEAT has reaching targets that it previously evolved. Even when reasonable reproductions are achieved, their CPPNs are significantly more complex than the original discoveries. Thus the implication is that evolving toward an objective creates a barrier to discovery by assuming that the stepping stones resemble the final objective.

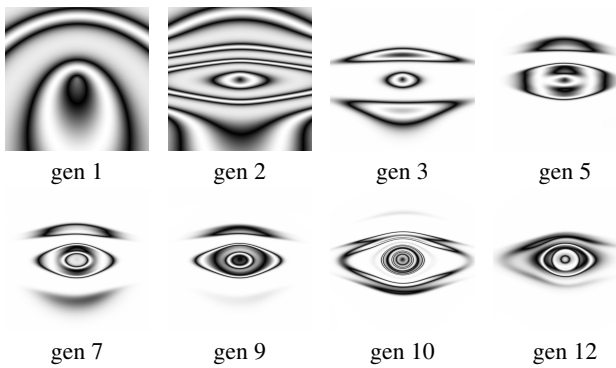


Figure 6: Stepping stones of non-objective image evolution. This sequence shows evolution guided by a single user with no explicit objective. The first image (gen 1) was selected from the initial population and progressed as shown to produce the published image (gen 12) known as the *Eye*, which emerged after 12 generations and is represented by a network of 10 functions and 16 connections. This example shows why the final objective is a poor heuristic for identifying stepping stones that lead to itself.

first seven generations or so, look nothing like the final eye at all. Given this observation, a "better" image-comparison metric would be even *worse* for this task because it would penalize the essential stepping stones (which do not look like the *Eye*) severely. Only in retrospect can we see why the stepping stones in figure 6 might lead to an eye. The user who originally found them chose them for *their own appeal*, not because they resemble the final published product. In fact, we can see the destructive effect of a "good" comparison metric (and see that the metric in equation 3 is actually good) in figure 7, which shows stepping stones in the sole 4,840-generation rediscovery of an *Eye*-like image. These stepping stones *do* resemble the final image, which is exactly why it takes so long to find it: Looking similar is exactly the *wrong* heuristic for identifying the most natural stepping stones.

In fact, the fallacy of the experimental-design objection exposes a fundamental flaw with objective-driven search in general: There is no a priori reason to believe that a metric that measures distance to the objective in *any* domain has a useful relationship to the essential stepping stones. The better the fitness function describes the objective, the more deceptive it may nevertheless be, which means that effort to better formalize the objective is misguided. The more ambitious and complicated the objective, the more profound this gap will be (as with the *Warp*, the *Butterfly*, and the *Skull*).

Yet while the more sensational result is the failure to reproduce the more complicated images, perhaps most sinister is what happens with the simpler images. They show that even the *appearance* of success is not genuine success. For example, although the *Crescent* superficially appears to have been rediscovered consistently (figure 5b), in *every* case the CPPN is at least twice as complex as the original representation (and usually three times more complex). The same is true for the *Simple* objective. Thus, *even if* the objective is reached because the problem is simple enough, a price will still be paid in the form of poor representation.

The reason for this representational inefficiency is that objective-based search by its nature encourages a *piecemeal* solution because it rewards small changes in parts of the solution that increase its resemblance to the objective, as can be seen in the unhealthy (yet successful) progression in figure 7. Each such small change requires adding a small amount of new structure to the representation to capture that piece. The result is a hodgepodge of functions that produce the correct pattern but in an inefficient way, which is why it takes thousands of generations to do what should take only a dozen.

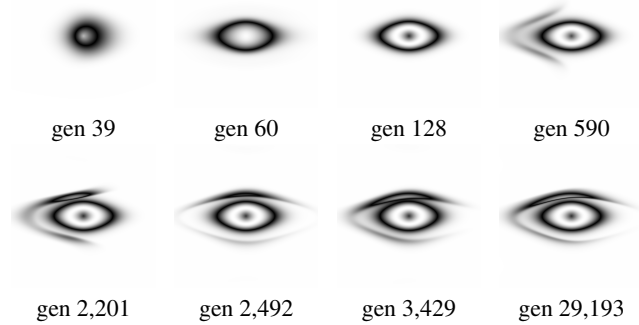


Figure 7: Evolving to an objective. The depicted sequence shows significant stepping stones reached by following the fitness gradient toward the *Eye* objective in figure 4c. Contrasting these stepping stones with those from the original discovery (figure 6) shows why resemblance to the target is not the right heuristic for identifying stepping stones for this problem.

The kind of *holistic* optimization that would have been ideal is impossible when the stepping stones that optimize the global structure do not resemble the final objective, as in figure 6.

In the short run, while we are only interested in immediate solutions to simple problems, this representational pathology may not matter, but in the long run, when we may want evolution to continue indefinitely or to build upon its past results, it will ultimately obstruct progress in the field. For that reason, the pathological effect of objectives on representation *even when search succeeds*, which has received little attention, merits significant further study. Many positive reported results may be unwittingly subject to this pathology, which is only uncovered in this study because the objectives were already evolved under different conditions.

This last point leads to a final important insight: *How* something is evolved matters. The mode of evolution impacts both the search and the representation, even when successful. Yet almost all our experiments are objectively driven. While the study in this paper focuses on images, the same principles apply to more practical domains as well, such as evolving controllers. For example, how do we know that the most natural stepping stones to a pole-balancing robot (which is a common benchmark [10,24]) are actually increasingly better at balancing a pole? In fact, the objective paradigm is so dominant that even considering the alternatives appears potentially radical. Yet we should not forget that the ultimate inspiration for our field is in nature, where evolution produced such marvels as photosynthesis, the flight of birds, and the human brain. Yet *none* of these discoveries were set as a priori objectives for the search. Instead, they are serendipitous discoveries on the road to nowhere in particular. Picbreeder [21,22] is similar; while its users stop at serendipitous waypoints on the road ahead, the system as a whole has no overriding purpose against which it is measured.² Is it a coincidence that its representations are so compact and its discoveries so rapid? Novelty search [14,15] is another example of a search process without an explicit a priori objective. Thus interactive evolutionary processes like Picbreeder and non-objective automated algorithms like novelty search emerge as possible alternatives to the current objective-driven paradigm that are deserving of further

²It is also interesting to note that users that come to Picbreeder with specific a priori objectives in mind often find the system frustrating [21]. Trying to re-evolve the skull from scratch (even as a human), which is virtually impossible, illustrates why discoveries on Picbreeder cannot simply be attributed to an uncanny (e.g. human) objective understanding of the search space.

investigation. In light of the results in this study, it is important to begin asking whether EC is inadvertently distorting a critical aspect of its original inspiration and thereby losing an essential ingredient: Nature has no final objective.

To conclude the discussion, while the discovery of e.g. the *Skull* could be dismissed by noting that NEAT only discovered it *once* but has shown no propensity to produce the same result again, perhaps our usual meaning of "result" is misguided. Perhaps the *result* in Picbreeder is not an individual image, or even a set of specific images, but rather the fact that it consistently produces interesting images in general (figure 1). We could dismiss the skull, but should we dismiss the skull, the car, the dolphin, the snake, the faces, the butterflies, the apple, the octopus, the insect, the pig, the planets, the tiger, the sunset, the candle, the eye, the penguin, etc. [21]? At what point do we accept that an EA is valuable not for its ability to produce a *particular* artifact that we want, but rather to consistently produce artifacts *in general*? In other words, is the traditional approach to *evaluating* evolutionary algorithms based on their ability to consistently achieve a particular objective flawed? After all, nature, the original inspiration for our field, has never discovered the very same organism in different lineages, yet its profusion of unparalleled discoveries remains no less potent as a clue to what is possible. What service then do we do to our understanding of evolutionary *algorithms* by judging them for their ability to repeat the same trick many times over? Is that really evolution's greatest trick, or is it a distraction from the real engine of evolution, which is creative discovery? Once, after all, is enough for nature.

7. CONCLUSIONS

The aim of this paper was to provide a unique perspective on both the downside of objectives and the cost to our field of judging EAs only for their ability to repeat the same achievement multiple times. While evolving without an objective is presently uncommon, such an approach allows fundamental principles (such as bilateral symmetry) to be discovered that may serve as stepping stones to interesting areas in the search space, which often could not have been reached by following a similarity metric. Experience with Picbreeder demonstrates the power of serendipitous discovery and highlights the importance of rewarding stepping stones for what they contribute in their own right. In this context, it may make more sense to judge EAs on their ability to produce a diversity of discoveries rather than on the probability of reaching the same objective over and over again. Perhaps then we may someday attain the unrealized potential of evolutionary algorithms to rival the achievements of nature.

8. REFERENCES

- [1] R. Dawkins. *The Blind Watchmaker*. Longman, Essex, U.K., 1986.
- [2] K. A. De Jong. *Evolutionary Computation: A Unified Perspective*. MIT Press, Cambridge, MA, 2002.
- [3] S. Droste, T. Jansen, and I. Wegener. On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science*, 276(1–2):51–81, 2002.
- [4] A. Eiben and J. Smith. *Introduction to evolutionary computing*. Springer-Verlag, 2003.
- [5] S. Fici and J. Pollack. Challenges in coevolutionary learning: Arms-race dynamics, open-endedness, and mediocre stable states. *Artificial life VI*, page 238, 1998.
- [6] D. Floreano, P. Dürri, and C. Mattiussi. Neuroevolution: from architectures to learning. *Evolutionary Intelligence*, 1:47–62, 2008.
- [7] D. B. Fogel. *Evolutionary computation: toward a new philosophy of machine intelligence, third edition*. Wiley-IEEE Press, 2006.
- [8] D. E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-wesley, 1989.
- [9] D. E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimization. In J. J. Grefenstette, editor, *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*, pages 41–49, Hillsdale, NJ, 1987. L. Erlbaum Associates Inc.
- [10] F. Gomez and R. Miikkulainen. Incremental evolution of complex general behavior. *Adaptive Behavior*, 5:317–342, 1997.
- [11] F. J. Gomez, J. Schmidhuber, and R. Miikkulainen. Efficient non-linear control through neuroevolution. In J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, editors, *Proceedings of the 17th European Conference on Machine Learning*, volume 4212 of *Lecture Notes in Computer Science*, pages 654–662. Springer, 2006.
- [12] D. James and P. Tucker. ANJI homepage. <http://anji.sourceforge.net>, 2004.
- [13] J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.
- [14] J. Lehman and K. O. Stanley. Exploiting open-endedness to solve problems through the search for novelty. In S. Bullock, J. Noble, R. Watson, and M. Bedau, editors, *Proceedings of the Eleventh International Conference on Artificial Life (Alife XI)*, Cambridge, MA, 2008. MIT Press.
- [15] J. Lehman and K. O. Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation*, 2011. To appear.
- [16] G. E. Liepins and M. D. Vose. Deceptiveness and genetic algorithm dynamics. Technical Report CONF-9007175-1, Oak Ridge National Lab., TN; Tennessee Univ., Knoxville, TN, 1990.
- [17] H. Mühlenbein and T. Manning. Convergence theory and applications of the Factorized Distribution Algorithm. *Journal of Computing and Information Technology*, 7(1):19–32, 1999.
- [18] M. Pelikan and D. Goldberg. Escaping hierarchical traps with competent genetic algorithms. Technical Report 2001003, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, 2001.
- [19] A. L. Ralescu. Generalization of the Hamming distance using fuzzy sets. Technical Report JSPS Senior Research Fellowship, Laboratory for Mathematical Neuroscience, The Brain Science Institute, RIKEN, Japan, 2003.
- [20] N. Saravanan and D. B. Fogel. Evolving neural control systems. *IEEE Expert*, pages 23–27, June 1995.
- [21] J. Secretan, N. Beato, D. B. D'Ambrosio, A. Rodriguez, A. Campbell, J. T. Folsom-Kovarik, and K. O. Stanley. Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary Computation*, 2011. To appear.
- [22] J. Secretan, N. Beato, D. B. D'Ambrosio, A. Rodriguez, A. Campbell, and K. O. Stanley. Picbreeder: Evolving pictures collaboratively online. In *CHI '08: Proceedings of the twenty-sixth annual SIGCHI conference on human factors in computing systems*, pages 1759–1768, New York, NY, 2008. ACM.
- [23] K. O. Stanley. Compositional pattern producing networks: A novel abstraction of development. *Genetic Programming and Evolvable Machines Special Issue on Developmental Systems*, 8(2):131–162, 2007.
- [24] K. O. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10:99–127, 2002.
- [25] K. O. Stanley and R. Miikkulainen. Competitive coevolution through evolutionary complexification. *Journal of Artificial Intelligence Research*, 21:63–100, 2004.
- [26] H. Takagi. Interactive evolutionary computation: Fusion of the capacities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.
- [27] S. Whiteson and P. Stone. Evolutionary function approximation for reinforcement learning. *Journal of Machine Learning Research*, 7:877–917, 2006.
- [28] N. Zaera, D. Cliff, and J. Bruten. (Not) evolving collective behaviours in synthetic fish. In P. Maes, M. J. Mataric, J. A. Meyer, J. Pollack, and S. W. Wilson, editors, *Proceedings of the fourth International Conference on Simulation of Adaptive Behavior (SAB'96)*, pages 635–642, Cambridge, Mass., 1996. MIT Press.