Leveraging Human Computation Markets for Interactive Evolution

Joel Lehman JOEL@CS.UTEXAS.EDU
Risto Miikkulainen RISTO@CS.UTEXAS.EDU
Department of Computer Science, The University of Texas at Austin, 2317 Speedway D9500, Austin, TX 78712

Abstract

Leveraging human input for selection in an evolutionary algorithm, i.e. interactive evolution, is effective when an appropriate domain fitness function is hard to quantify, but where solution quality is easily recognizable by humans. However, single-user applications of interactive evolution are limited by *user fatigue*: Humans become bored with monotonous fitness evaluations. This paper shows that user fatigue can potentially be bypassed through human computation markets that enable directly paying for human input. Experiments evolving images show that purchased human input can be leveraged more economically when evolution is seeded with products from a purely-computational aesthetic measure. Further experiments in the same domain validate a system feature, demonstrating how human computation can help guide interactive evolution system design. Finally, experiments in an image composition domain show how the approach can facilitate large-scale interactive evolution in tasks that are not inherently enjoyable. In this way, combining human computation markets with interactive evolution facilitates mechanical application of a powerful form of selection pressure.

1. Introduction

A critical component of any evolutionary computation (EC) experiment is selection, i.e. how the parents of the next generation are chosen from the current population. Successfully applying an EA to a given domain often requires choosing an appropriate *fitness function* to guide search. However, intuitive choices for fitness functions may often fail to identify the intermediate steps that lead to the solution (Lehman & Stanley, 2011; Goldberg, 1987), and some concepts intuitive to humans remain difficult to quantify algorithmically (Secretan et al., 2012; Takagi, 2001).

In such cases, one way to bypass these difficulties is through interactive evolutionary computation (IEC; Takagi, 2001), wherein humans act as a fitness function, actively selecting which solutions to evolve further. The insight is that humans may be able to evaluate a characteristic even when it cannot be mechanically recognized.

However, a significant problem in IEC is user fatigue: A single user can only perform so many evaluations before becoming tired or bored (Takagi, 2001). A recent solution to this problem is to create collaborative IEC websites whereby without financial incentive users cooperate to evolve complex artifacts they could not have evolved alone (Secretan et al., 2012; Clune & Lipson, 2011; MacCallum et al., 2012).

This approach is promising when task domains are designed to be enjoyable, e.g. creative domains like open-ended image, shape, or music evolution (Secretan et al., 2012; Clune & Lipson, 2011; MacCallum et al., 2012). However, when attempting to apply the approach to arbitrary domains there are two significant limitations: (1) sustained evolution depends upon the task domain being engaging enough to continually attract many volunteer users, and (2) implementing the idea requires creating the non-trivial system architecture that composes a collaborative evolution website.

An interesting potential solution to these problems is provided by recently developed human computation markets (HCMs). In these markets it is possible to pay for human input in arbitrary tasks and thereby keep humans motivated even when the task is not engaging. Thus this paper explores such an approach called HCM+IEC that uses HCMs to perform selection in an IEC.

More specifically, it focuses on three ideas: First, even

ICML Workshop: Machine Learning Meets Crowdsourcing, Atlanta, Georgia, USA, 2013.

if the domain to be used with IEC is itself fun (e.g. evolving aesthetic images), IEC websites face the bootstrapping problem common to all user-generated content sites. That is, at such a site's launch, when attracting users is most important, the site is *least* engaging due to lack of content. Thus the first contribution of this paper is to suggest that markets for human computation can help overcome this bootstrap problem: Initially users can be paid to generate content. For this reason, experiments with such an aim apply IEC+HCM in an image evolution domain. The results show that human computation can be more efficiently leveraged if a computational aesthetic measure (Lehman & Stanley, 2012) first algorithmically generates an interesting diversity of images upon which humans can further elaborate.

Second, when designing an IEC website or a single-user IEC system, often many design decisions about the underlying algorithm must be made that will significantly impact the quality of the system's output. However, such important decisions often are guided only by the preferences of the system designers. The second contribution of this paper is thus to suggest that IEC+HCM can be applied to conduct controlled experiments that measure the impact of a design decision on the quality of an IEC system's products. Experiments in the same image evolution domain show that removing a well-motivated feature results in measurably less aesthetically pleasing pictures, demonstrating the potential for IEC+HCM to facilitate principled IEC system design.

Third, there are EC problems that could benefit from large-scale human selection but for which a collaborative IEC website will not likely be an appropriate solution. That is, most current IEC websites rely on self-directed users to produce content, and such content is produced irregularly and only to the extent that volunteer users *enjoy* evaluating artifacts in the domain. Thus the third contribution of this paper is to demonstrate how IEC+HCM can be used instead of a collaborative IEC website in one such condition, i.e. when the task domain is not enjoyable.

The conclusion is that HCMs offer a mechanism for converting money into a powerful form of selection pressure that may prove a productive tool for interactive evolution.

2. Background

In this section, the foundational technologies applied in the experiments in this paper are reviewed.

2.1. Interactive Evolution

Applying human judgment to perform selection in an evolutionary algorithm is called interactive evolutionary computation and is motivated by the difficulty in quantifying intuitive concepts that are readily recognized by humans (e.g. aesthetic appeal). While IEC has been explored in the context of single-user applications (Takagi, 2001) and collaborative websites (Secretan et al., 2012; Clune & Lipson, 2011), it has only been superficially explored in the context of HCMs (Chou et al., 2012).

Previous studies with IEC have demonstrated its promise for evolving complex structures (Gruau & Quatramaran, 1997; Woolley & Stanley, 2012). A representative example is the Picbreeder website (Secretan et al., 2012) that facilitates indirect user collaboration to evolve aesthetic images. On the site, users can discover, rate, and extend previously evolved images that are represented by compositional pattern producing networks (CPPNs; Stanley, 2007). Similarly designed websites may be one general path to large-scale IEC and compelling evolved artifacts.

However, evolution in such websites is typically undirected (i.e. driven by users' whims on what to create) and public (i.e. driven by users' ability to discover and elaborate upon existing content); for commercial IEC applications the ability to more directly guide the evolutionary process may be important and additionally it may be necessary for evolved content to be kept more private (i.e. not stored such that all content is publicly accessible).

These limitations motivate exploring new approaches for large-scale IEC. A promising resource that can be leveraged for such purposes is human computation, which is reviewed next.

2.2. Human Computation

While the range of tasks solvable by computers continues to expand, there remain tasks that are challenging to solve computationally but are trivial for humans to solve. Examples of such tasks are recognizing written text (Von Ahn et al., 2008), identifying objects in images (Von Ahn, 2006), or evaluating aesthetic appeal (Secretan et al., 2012; McCormack, 2005). As a result, it may be useful to leverage *human computation* (Von Ahn et al., 2008) to automatically integrate human insight into algorithmic processes. Such human computation can often be made more scalable by employing *crowdsourcing* (Kittur et al., 2008; Orkin & Roy, 2007), whereby many small contributions from a diffuse group of people (often online) are aggregated.

For example, while CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) separate humans from machines by generating tasks that are easily solvable by humans but difficult for machines, the widely-deployed reCAPTCHA system (Von Ahn et al., 2008) acts as a CAPTCHA while at the same time leveraging human computation to transcribe words from old books. Similarly, "games with a purpose" (GWAP; Von Ahn, 2006) are designed such that human enjoyment results from deriving and verifying solutions to problems that are not yet solvable computationally.

ReCAPTCHA and GWAPs show that sometimes users can be enticed to generate useful computation without economic incentive. However, it is unclear how to transform an arbitrary human computation task into a fun or necessary process such that the task's solution is a byproduct. Thus it may be simpler or cheaper sometimes to simply *pay* a human to perform the desired task through a HCM. The most well-known such marketplace is the Amazon Mechanical Turk (AMT; Ipeirotis, 2010), which is the system used in the experiments in this paper.

AMT exposes an interface to programmers that allows them to upload human intelligence tasks (HITs) which specify a desired task, an interface for humans to perform it, and the monetary reward for successfully completing the task. Once a human completes the HIT, the results can be queried and approved so that the human user can be paid. In this way, HCMs like AMT allow seamless integration of algorithms with arbitrary human input through economic exchange.

3. Approach

While previous approaches to IEC are limited by user fatigue or require that a domain be enjoyable to attract users, the approach in this paper, IEC+HCM, avoids such issues by paying users for performing IEC evaluations through a HCM (figure 1). Of course, the trade-off is that with IEC+HCM there is an explicit economic cost for each evaluation.

The particular HCM applied in the experiments described here is AMT. Recall that AMT provides a computational interface for posting small computational tasks with a set monetary reward which must be high enough to entice workers. Importantly, the AMT interface can be applied to automate IEC tasks (e.g. by presenting a user with an artifact or behavior and querying for evaluation via a web form). Thus, tasks can be mechanically created and uploaded to AMT, results can be collected, and participants can be paid in real time. In this way the methodology can scale to arbitrary limits given enough money and available users in the human computation market, potentially overcoming previous limitations to easily implementing IEC in any domain on a large scale. (However, note that this paper is an initial small-scale exploration of the feasibility of such ideas; the experiments presented here involve only 272 different AMT users in total.)

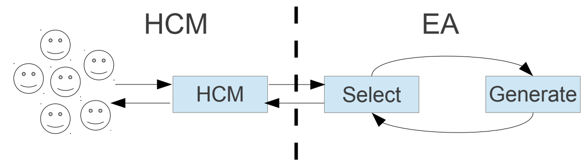


Figure 1. **The IEC+HCM approach.** During selection, the EA uploads evaluation tasks to the HCM, and they are then completed by human users. The results are applied by the EA to create the next generation. In this way, an EA can be driven by the human judgment of many non-experts.

Importantly, there are many potential ways to combine AMT (or other HCMs) with an IEC algorithm. One design decision is how tasks should be divided. For example, a task sent to a HCM for completion could consist of evaluating only a single artifact, evaluating the evolutionary algorithm's entire population, or guiding multiple generations of IEC evolution. In this paper, tasks were divided into evaluations of an entire population, similarly to how a user influences a single generation of evolution in most single-user IEC applications (Takagi, 2001).

Another design decision is what type of input should be gathered from human users; such input could consist of only which artifact in a population was most preferred, or could require individually rating each artifact. Individual ratings were gathered in this paper to enable comparisons between generations and runs, and to encourage greater deliberation during evaluation.

A final decision is how user evaluations of artifacts guide evolution. In the approach in this paper each user evaluates only one generation at a time, and multiple independent evaluations of the same population are combined together to allocate offspring for the next generation. In particular, children are allocated to artifacts in proportion to how many users rate them most-highly.

Thus while other approaches to IEC+HCM may also be viable, the described approach reasonably combines IEC and HCM, and its design decisions form a coherent methodology.

4. Experiments

In following sections, experiments are presented that apply IEC+HCM to two domains. Results in evolving aesthetic images are first presented as a domain characteristic of those appropriate for collaborative IEC websites (Secretan et al., 2012). The second domain evolves compositions of image layouts as an exemplar for where economic incentives are crucial for success.

4.1. General Experimental Setup

For all experiments, AMT was used to buy human computation and a set price of $0.05 USD was paid per user completion of a task. A standard genetic algorithm was applied with a small population size (nine individuals) characteristic of many IEC domains (Takagi, 2001). All runs consisted of ten generations, with one task uploaded per generation. Three runs of each method were performed, providing a qualitative proof-of-concept of the kind of investigations that IEC+HCM enables.

Tasks uploaded to AMT contained nine images, i.e. the entire population, and required users to rate the images' aesthetic appeal on a scale from one to five (where five is the best). Because aesthetic judgment is subjective and varies between individuals, each task was evaluated by five separate AMT users to get a more representative sample. In particular, images were selected proportionally to how many users rated them most highly among the nine presented (i.e. only a user's highest-rated images would contribute to selection).

4.2. Evolving Aesthetic Images with IEC+HCM

The first two experiments explored an image evolution domain that implements an encoding similar to the Picbreeder collaborative IEC website (Secretan et al., 2012) and explored elsewhere in single-user IEC applications (Stanley, 2007). In particular, in these systems images are represented by ANN-like networks called CPPNs, which are briefly reviewed in the next paragraph (a more detailed introduction is given by Stanley 2007).

Importantly for the second experiment in this domain, while each node in a traditional ANN has a sigmoidal activation function, the nodes of CPPNs each have an activation function selected from a set of such functions. The choice of possible activation functions in the set is motivated by potential regularities they can induce when the CPPN is queried. For example, sinusoidal functions may be included to induce repetition and Gaussian functions may be included to induce symmetry. In this domain, a CPPN is mapped to the image it represents in the following way: For each pixel in an image, the CPPN's inputs are set to its scaled Cartesian coordinates, and the output of the network is interpreted as a grayscale pixel value. In effect, the CPPN thus represents a pattern over a coordinate space, which in this case is interpreted as a picture.

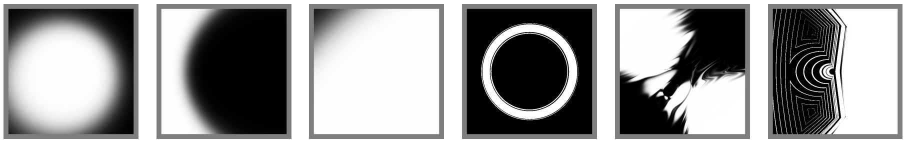
4.2.1. Bootstrapping Collaborative IEC

The goal of this experiment is to show that IEC+HCM can be applied to evolve aesthetic images through selection from a diffuse cloud of paid users. A practical application of such a technique is to *bootstrap* newly launched collaborative IEC websites with initial content. That is, to make a site more engaging, users can be initially paid to evolve content.

However, because IEC+HCM incurs a financial cost for each evaluation, it becomes important to leverage human input as efficiently as possible. Thus a promising approach may be first to generate a diversity of content algorithmically that is more appealing than the random genomes that would otherwise seed IEC+HCM. One such approach, which is applied here, is to evolve seed artifacts using a computational measure of impressiveness (Lehman & Stanley, 2012). Examples of such evolved artifacts and how they differ from randomly generated artifacts are shown in figure 2. Note that other automated aesthetic generators could have been applied, what is important is that although not perfectly aligned with human taste, such impressive artifacts are more engaging than randomly generated ones.

To investigate whether such seeding is useful, two versions of IEC+HCM are run: one method that is first seeded with pre-evolved impressive images, and another that is initialized with random genomes. For the unseeded runs, evolution starts from simple random CPPNs in the same way as most other CPPN-encoded image evolution applications (Secretan et al., 2012; Stanley, 2007). For the seeded runs, the setup of Lehman & Stanley (2012) was applied to first evolve impressive artifacts, of which the most impressive CPPNs from 20 separate runs were sampled to seed evolution. Figure 3 shows the products of both methods.

Because judging aesthetic appeal requires subjective human evaluation, AMT was also applied to investigate the products of the two methods. In particular, the images from figure 3 (excluding the initial unseeded images) were placed in random order and



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

Figure 2. **Comparing random and impressive images.** The images shown in (a),(b), and (c) are representative of images generated by random genomes in the image evolution domain, while the images shown in (d),(e), and (f) are examples of images evolved through the impressiveness metrics. Importantly, the impressive images differ qualitatively and are noticeably more complex than the random images.

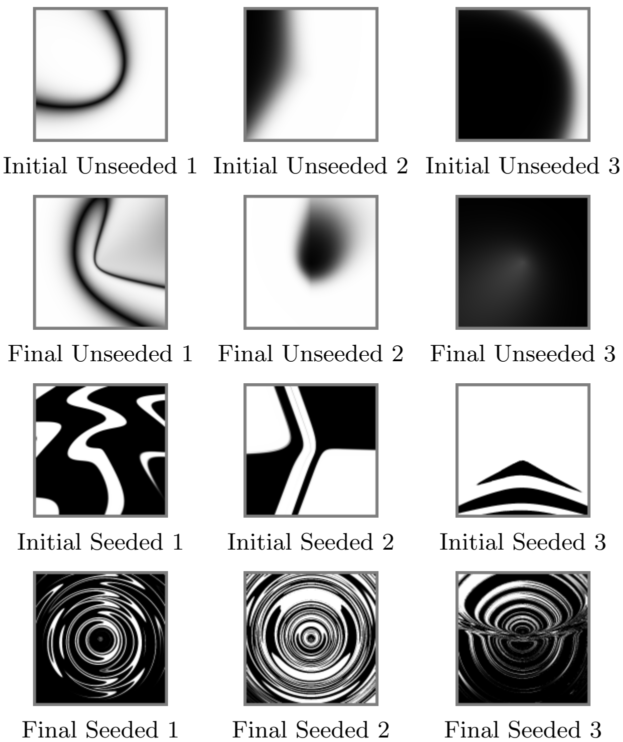


Figure 3. **Products of the Seeding Experiment.** Images from the three runs of the seeded and unseeded methods. The labels indicate whether the images are the most-preferred images from the initial or the final generation, whether they are from the seeded or unseeded method, and finally during what number run (out of three) they are generated. The main results are (1) that there is a large difference in complexity and quality between the unseeded and seeded runs, and (2) that for the seeded runs there is a noticeable divergence between the initially-preferred seed and the final most-preferred evolved image.

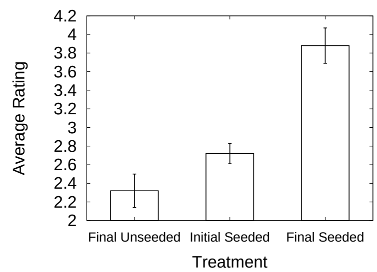


Figure 4. **Seeding experiment evaluation.** An independent evaluation comparing the champions of the final generation of the unseeded method (final unseeded), the initial generation of the seeded method (initial seeded), and the final generation of the seeded method (final seeded) are shown. The main result is that the final seeded champions are on average rated significantly more aesthetically pleasing than both the final unseeded and initial seeded images.

uploaded to the same AMT evaluation task used for IEC, but with a larger number of separate user evaluations (20 instead of five). The results of this evaluation (seen in figure 4) show that the champions from the *first* generation of the seeded runs (i.e. the most preferred "impressive" seed image) are rated significantly more aesthetically pleasing than are the *final* generation champions from the unseeded runs (Student's t-test; $p < 0.05$). That is, on average users preferred at least one of the pre-evolved seed images to the final products of the unseeded runs. Furthermore, the champion of the final generation of the seeded runs (i.e. the most-preferred product of human elaboration of the seed images) is rated significantly more pleasing than both the initial generation of the seeded runs and the final generation of the unseeded runs (Student's t-test; $p < 0.05$). In this way, the results support the hypotheses that IEC+HCM can be leveraged to evolve increasingly aesthetically pleasing artifacts and that seeding IEC+HCM with pre-evolved artifacts can more efficiently leverage human evaluations. Thus seeded IEC+HCM may be a viable technique for bootstrapping collaborative IEC websites.

4.2.2. Validating Components of an IEC System

The next experiment is motivated by the desire to make principled design decisions while creating a collaborative IEC website or single-user IEC application. That is, it is difficult for a system designer to decide objectively on appropriate parameter settings or feature encodings, especially when the quality of such decisions depends upon subjective factors aggregated across all targeted users (e.g. the aesthetic quality of artifacts under such a decision).

While single-user IEC applications are easy to revise from user feedback even after they have been first released, launching an IEC website inherently involves a certain level of commitment to the domain. That is, changing the domain after the website has launched may invalidate already-evolved content, potentially alienating users whose creative products are deleted. Therefore it is desirable to avoid such problems and launch a better initial product.

A potential solution is to run controlled experiments with IEC+HCM to collect empirical evidence of a change's impact from a representative sample of potential users. That is, the quality of results from IEC with different parameter settings or features can be compared, by paying users through AMT to perform selection and then by paying other users to compare the final results.

Thus as a simple example, the second experiment investigates the claim that the additional activation functions of CPPNs improve the aesthetic quality of CPPN-evolved images beyond the use of simpler ANNs (Stanley, 2007; Secretan et al., 2012). To investigate this idea, a third version of the image evolution task was devised, but with simple ANNs (i.e. standard ANNs with only a single sigmoidal activation function) instead of CPPNs (which have an extended set of activation functions). In this way, the aesthetic quality of products evolved with CPPNs could be compared to those evolved with simpler ANNs. Furthermore, taking into account the advantages of seeded IEC+HCM runs demonstrated in the previous experiment, only a seeded method with simple ANNs was considered.

The effect of replacing CPPNs with ANNs on the results of IEC+HCM is shown in figure 5; as expected, these images differ noticeably from the previous results with CPPNs shown in figure 3. An empirical investigation of the aesthetic difference between the seeded IEC+HCM methods with CPPN and ANNs was then conducted similarly to the previous experiment: AMT users compare the products of this simple ANN method (shown in figure 5) with those of the previous IEC+HCM method with CPPNs (i.e. the final seeded images from figure 3). Figure 6 shows the results: Expanding the set of activation functions in CPPNs facilitates evolving more aesthetically pleasing images. While not surprising, this intuitive result demonstrates how IEC+HCM may generally be used to investigate the impact of different feature encodings.

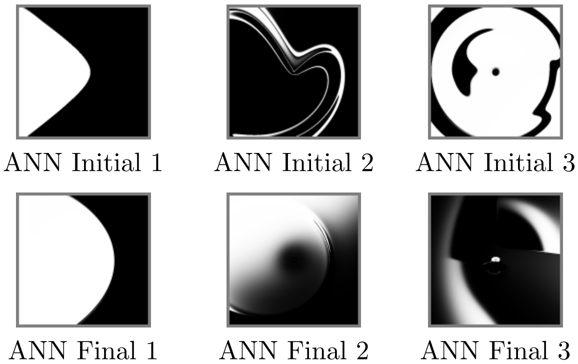


Figure 5. **Products of the Seeded ANN Runs.** Images are shown from the three runs of seeded IEC+HCM with ANNs (instead of with CPPNs as in the previous experiment). There is a large divergence between the initial seed and final image in two out of three runs. The qualitative difference between these images and those evolved with CPPNs (figure 3) suggests that the added activation functions of CPPNs impact the kind of images likely to be evolved.

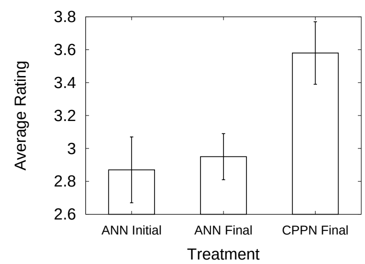


Figure 6. **Feature Validation Experiment Evaluation.** An independent evaluation comparing the initial generation and final generation champions of the seeded method with ANNs (ANN initial and ANN final) with the final generation champions of the seeded method with CPPNs (CPPN final) are shown. The main result is that the final generation CPPN images are judged significantly more aesthetically pleasing than either of the two classes of ANN images (Student's t-test; $p < 0.05$). The conclusion is that CPPNs facilitate evolving more aesthetic images than ANNs.

4.3. Evolving Image Layouts with IEC+HCM

The third experiment investigates whether IEC+HCM can expand the range of domains where large-scale IEC can be effectively applied. While it is currently applicable only to domains that are sufficiently enjoyable to attract volunteer users, the IEC+HCM approach can potentially be applied to any domain regardless of how engaging it is, and can be scaled to the extent that funds are available to do so. Of course, the more painful domain evaluations are, the more it may cost.

Thus to examine whether IEC+HCM can be applied to domains not inherently enjoyable, the third exper-

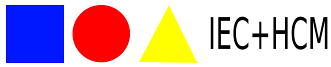


Figure 7. **The Image Layout Domain.** The image layout experiment evolves a composition of the four shown images through IEC+HCM. The encoding is a simple list of Cartesian coordinates that specify the offset of each of each image. Mutation perturbs the coordinates of one image out of the four, adding to the x and y coordinate a separately number chosen uniformly between -50 and 50.

iment explores an intuitively less enjoyable task, that of evolving the layout of an image composition. In particular, the task is to evolve the relative positions of a fixed set of images (seen in figure 7) to maximize the aesthetic appeal of the composition. Unlike the image evolution domain, the potential for novelty is limited because the components of the image are always the same and uninteresting.

The specifics of the domain and encoding are illustrated by figure 7. Note that the same IEC+HCM setup as in the previous experiments was adapted for this third experiment, utilizing only a single unseeded method. While seeding with impressive pre-evolved layouts might accelerate progress in this domain, such seeding is not necessary to verify the hypothesis.

The products of this experiment are shown in figure 8. The results were validated similarly to the previous experiments, by randomizing the images in figure 8 and presenting them to be rated by a larger set of AMT users. However, instead of comparing between methods, the comparison of evolved artifacts is over generations. The idea is to demonstrate that progress in aesthetic evolution is occurring. The aggregated ratings from the larger validation evaluation are shown in figure 9. As expected, the most-preferred layouts from the final generation are rated significantly more pleasing in appearance than those from the first generation, thus supporting the conclusion that evolutionary progress was facilitated by IEC+HCM in this domain. Note that while the domain itself is somewhat trivial, the results provide an existence proof that IEC+HCM can extend large-scale IEC to domains that are not inherently fun.

5. Discussion and Future Work

This paper investigated the idea of leveraging markets for human computation to support large-scale IEC in three ways. Exploratory experiments in this paper showcase how the ability to pay for human computation potentially can bootstrap IEC websites, inform the design of such websites or single-user IEC systems,

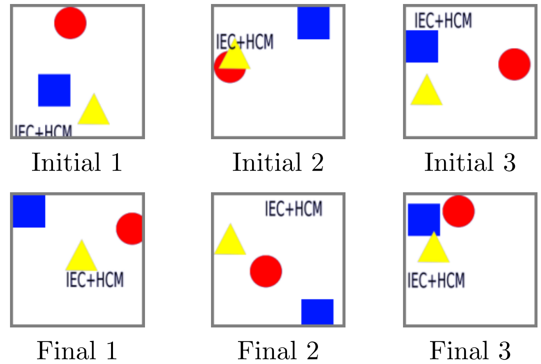


Figure 8. **Products of the Layout Evolution Experiments.** Images are shown from the three runs of IEC+HCM in the layout evolution domain. In particular, the most-preferred image from the initial and final generation of the runs are shown. Over evolution, the images composing the layouts expand to better fill the space and partially-obscured text becomes readable. The conclusion is that IEC+HCM can be successfully applied even in domains that are not inherently enjoyable.

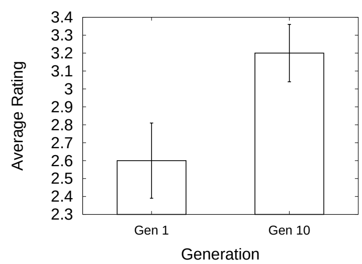


Figure 9. **Image Layout Experiment Evaluation.** An independent evaluation comparing the champions of the first and final (tenth) generations of IEC+HCM runs in the image layout domain is shown. The main result is that the image layout of final champions is judged significantly more aesthetic than that of the first generation (Student's t-test; $p < 0.05$).

and can act as a viable alternative to such websites when the domain is not inherently enjoyable.

In this way, an interesting advantage of the IEC+HCM approach is that it bypasses the significant problem of user fatigue in IEC (Takagi, 2001) without constraining the domain. Of course, the trade-off is that pairing IEC with human computation incurs an explicit financial cost per evaluation. Thus large-scale IEC+HCM may be most applicable for unengaging domains limited by difficulty in applying appropriate selection pressure, and also possibly for commercial applications where the cost of IEC+HCM is less than the value of the evolved artifact.

So while the IEC+HCM mechanism can be leveraged to improve the design and engagement of single-user IEC systems and collaborative IEC websites, its most interesting implication may be that exploiting it on a large scale may potentially lead to results exceeding current approaches in evolutionary robotics or artificial life. That is, to the extent that current approaches are limited by lack of appropriate selection pressure (Zaera et al., 1996; Miconi & Channon, 2006; Lehman & Stanley, 2011), and to the extent that human judgment can remedy such limitations (Gruau & Quatramaran, 1997; Woolley & Stanley, 2012), human computation may be a technique that can be exploited to further the state of the art in EC. For example, large-scale IEC+HCM with a significant budget applied to evolving virtual creatures might produce creatures with complexity and functionality beyond the reach of current methods. In this way, an interesting direction for further experimentation is to apply IEC+HCM to evolve controllers for evolutionary robotics or artificial life experiments.

6. Conclusion

This paper explored combining interactive evolution with human computation markets to purchase a powerful form of selection pressure. The preliminary promise of the approach was shown in experiments evolving aesthetic images and the layout of image compositions. Applying the same techniques in other domains limited by lack of appropriate selection pressure may enable evolution of more complex artifacts or behaviors than previously possible. The conclusion is that human computation markets may be an important tool for supporting collaborative IEC websites as well as for extending the reach of large-scale IEC beyond only task domains that are enjoyable.

References

Chou, C., Kimbrough, S., Sullivan-Fedock, J., Woodard, C.J., and Murphy, F.H. Using interactive evolutionary computation (iec) with validated surrogate fitness functions for redistricting. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference*, pp. 1071–1078. ACM, 2012.

Clune, J. and Lipson, H. Evolving three-dimensional objects with a generative encoding inspired by developmental biology. In *Proceedings of the European Conference on Artificial Life*, 2011.

Goldberg, David E. Simple genetic algorithms and the minimal deceptive problem. In Davis, L. D. (ed.), *Genetic Algorithms and Simulated Annealing, Re- search Notes in Artificial Intelligence*. Morgan Kaufmann, 1987.

Gruau, F. and Quatramaran, K. Cellular encoding for interactive evolutionary robotics. In *Fourth European Conference on Artificial Life*, pp. 368–377. MIT Press, 1997.

Ipeirotis, P.G. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21, 2010.

Kittur, Aniket, Chi, Ed H, and Suh, Bongwon. Crowdsourcing user studies with mechanical turk. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 453–456. ACM, 2008.

Lehman, Joel and Stanley, Kenneth O. Abandoning objectives: Evolution through the search for novelty alone. *Evol. Comp.*, 19(2):189–223, 2011.

Lehman, Joel and Stanley, Kenneth O. Beyond open-endedness: Quantifying impressiveness. In *Proceedings of Artificial Life Thirteen (ALIFE XIII)*, 2012.

MacCallum, R.M., Mauch, M., Burt, A., and Leroi, A.M. Evolution of music by public choice. *Proceedings of the National Academy of Sciences*, 109(30):12081–12086, 2012.

McCormack, J. Open problems in evolutionary music and art. *Applications of Evolutionary Computing*, pp. 428–436, 2005.

Miconi, Thomas and Channon, Alastair. An improved system for artificial creatures evolution. In *Proceedings of the Tenth International Conference on Artificial Life (ALIFE X)*, pp. 255–261. MIT Press, 2006.

Orkin, Jeff and Roy, Deb. The restaurant game: Learning social behavior and language from thousands of players online. *Journal of Game Development*, 3(1):39–60, 2007.

Secretan, J., Beato, N., D'Ambrosio, D.B., Rodriguez, A., Campbell, A., Folsom-Kovarik, J.T., and Stanley, K.O. Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evol. Comp.*, 2012. ISSN 1063-6560.

Stanley, K.O. Compositional pattern producing networks: A novel abstraction of development. *Genetic Programming and Evolvable Machines*, 8(2):131–162, 2007. ISSN 1389-2576.

Takagi, H. Interactive evolutionary computation: Fusion of the capacities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.

Von Ahn, L. Games with a purpose. *Computer*, 39(6): 92–94, 2006.

Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008.

Woolley, Brian G. and Stanley, Kenneth O. Exploring promising stepping stones by combining novelty search with interactive evolution. *CoRR*, abs/1207.6682, 2012.

Zaera, N., Cliff, D., and Bruten, J. (Not) evolving collective behaviours in synthetic fish. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. MIT Press Bradford Books., 1996.