Human Tended Gardens of Evolutionary Design

Matthew Hockenberry, Ernesto Arroyo

Creative Synthesis Collaborative, Universitat Pompeu Fabra
167A Erie St., Cambridge MA 02139
hock@creativesynthesis.net, ernesto.arroyo@creativesynthesis.net

Abstract

This paper describes an artificial creative system that simulates basic creative design behavior through the use of pseudo-genetic design supplemented by human recognition and evaluation. While it remains unclear if the system is truly creative itself, it provides the necessary support structure for a design platform that reshapes creative decision making as a question of design growth rather than manufacture.

Introduction

Design is a difficult problem space. Identifying the 'best' design solution for a particular problem is a nontrivial, often difficult process. Furthermore, the heuristics for evaluating design vary considerably depending on the domain, the task for the design object and the user who will make use of the design. At its most base level, a design is just a plan. The implementation of such a design is a designed object, or more commonly just an object. For our purposes we consider both physical objects (like watches) and virtual objects (like your computer's system time display). The design describes the aesthetic values of the object (what it looks like) and the functional values (what it does with regard to a given stimulus).

Any understanding of what a 'best' design is must necessarily incorporate both the aesthetic and functional values and the resultant impression of the object's use - is an object useful, given a purpose for a kind of person [Doubleday et al., 1997]? This is the concept of the object's affordance. A watch possesses some affordance for telling time and also some affordance for hitting nails. One may argue that the designed purpose of the watch is to tell time - and most evaluations of the object would bare this out. Watches are generally made of precocious metals, with fragile parts. This doesn't lend itself as readily to hitting nails as a hammer, partly because of human perception and partly because of physical constraints. The affordance of the object must incorporate both. Affordance is a set of action possibilities - or perceived action possibilities - for a particular object [Gibson, 1977]. Intuitively a functionally designed object is the one that produces the tightest perceived action possibilities consistent with the designer's intent.

Aesthetics plays a dominant role in this discussion as well. Good design is often associated with some kind of positive aesthetic judgment. A watch made of gold may be less functional in some respects, than a watch made of platinum - but the gold watch hold's cultural and personal value. These judgments that affect our conception of what good design is. However, although aesthetics are important in evaluating design, it may be convenient to consider them as another functional property. Design, unlike art, is not entirely concerned with aesthetic judgments. We can quickly draw to mind examples of 'good' design that don't seem to require aesthetic appreciation. Aesthetics, for our purposes, could be incorporated into the concept of affordance. For an object to afford any use it all, it must be aesthetically palatable enough for someone to pick it up in the first place. The value we associate with design seems to be inextricably linked to human experiential use and perception.

The Fitness of Design

A lot of this still seems fairly arbitrary. Part of the problem is what we mean by good. Do we mean usable? Aesthetically pleasing? Gets the most adsense revenue? The way we define ‘good’ is essentially the fitness of a particular design. In life a ‘good’ life form is simply one that can continue to live and, ultimately, reproduce. To live a life form needs to be successful at getting nutrients (eating and breathing) and in reproducing to continue its genetic code. In the real world a flower that is incredibly beautiful, but extremely difficult to reproduce, fails the real world test of fitness.

In a computer simulation we can be a little more flexible with the rules of the world. Let’s imagine that flowers could somehow gain nutrition from human attention - if human finds the flower aesthetically interesting, functionally valuable, or just fun to play with it receives sustenance. If this was true it stands to reason that the most beautiful flower and the most functional flower might be considered the most fit. They would continue to grow and reproduce because they got more attention and consequently received more nutrition than other flowers.

Human attention is an interesting and fickle thing. We could imagine that as the flowers reproduce they become more and more populous. Humans get used to them and look at them less. Eventually, even though it was once thought beautiful and attractive, our aesthetically pleasing flower comes to be considered common and ugly. The plant dies out. Our functional flower, on the other hand, would continue to reproduce until a more functional competitor displaces it.

There are non-intuitive solutions here as well. Another beautiful flower that has evolved with a similar nutrient mechanism (feeds on attention) but limits its reproduction could be more viable. It will always remain scarce and never meet the fate of its cousin. This is a dangerous reproduction strategy in the real world, but not in our simulation.

Design is a creativity problem [Dorst and Cross, 2001]. Finding a 'great' design is not necessarily a purely mechanical affair. Despite this, developing a design so that it is 'good' – at least functionally so, could be mechanical with a process of iterative evaluation.

Design as a Search Problem

We'll begin now by describing design as a search problem, where the goal is to find the best design is one that meets our test of fitness in terms of human attention. This supports both a functional and aesthetic evaluation. Genetic algorithms are commonly used to identify the exact or approximate solutions to optimization and search problems. These approaches are often categorized such that there is an implied assumption of a 'best solution' that can be mathematically proven. For example, a search solution can be shown to better than a previous solution if it is faster than the previous solution. The best solution, where best is defined by computation time, is simple to identify.

It's true that not all test of fitness are simple mathematical propositions. The 'best' search might also require some human understanding. The casual notion of best search is the one that produces the result one would most prefer. That preference may be difficult to describe - but it is not impossible. Still, there is the implicit notion that there is one best answer to the problem.

The problem space we are considering, design, arguably lacks the notion that there is one best answer. A large part of this is that the problem space is made up of search criteria that are deeply rooted in experiential perceptions of human beings. Consider the criteria with which we might evaluate a successful design. Aesthetics? Aesthetics have been notoriously difficult to codify. Ergonomics? Understanding of comfort varies depending on person, and (demonstrably in Japanese factory workers) task. Quality? While we might be able to rank materials and construction methods, this remains somewhat subjective. Economics? Identifying that something is 'a good value' remains a problem in economics. Function? Evaluating function varies considerably depending on the kind of object.

There is almost no criterion with which we might evaluate design that is not in some part, subjective. In fact, the evaluation of design seems to springboard us into complex problems in a large number of domains. Our solution to this problem has been to record the human experience with the design with the understanding that it requires subjective interpretation to be meaningful.

Collecting Accounts of Human Experience

Recycled Research is our experiment in design, experimental methodology, sharing and process - it was designed to address a number of problems that affect the design research and software engineering communities. The essential intuition of Recycled Research is to take advantage of a decentralized software based experimental system to store human (empirical) experience with interactions and responses to stimuli. We target recycled research to design based research inquiries.

If we choose to define design in terms of human attention and perception, this kind of system seems well suited. Human experiential accounts of design interaction are valuable to us as tools for measuring design. We can't produce them on our own, we require actual understanding of how people interact with software, interpret stimuli, or develop skills. There are numerous examples of software design that are out there, interacting with humans, which have the unfortunate lack of a design researcher watching them. Recycled Research codifies these human experiential accounts of interaction.

Recycled Research is the coupling a structured research goal to a distributed medium, such as software, for the purposes of structured data collection and hypothesis testing. The idea that we can couple a useful software tool with an experimental mechanism is happening anyway, every piece of software is conducting at least one experiment – "can the user understand the interface well enough to get something done." This extends beyond the value found in current trends of data collection for usability and pushes the direction into external goals [Couper, 2000]. Procedurally, first-tier users, which we think of as something like assistant researchers or experimenters, install software on websites, installations, or intranet where it will be exposed to second-tier users, those we generally think of as participants in traditional studies and traditionally describe as end users - these are the people we are evaluating the design against.

We have targeted the web as the principle mechanism for enabling this kind of interaction. This is limiting, because it doesn't allow us to evaluate certain kinds of three-dimensional designs (at least not well). However, it does capture print, web and video design, which is not insignificant.

Garden Tending

If we think of this design research as the garden mentioned in the title, the goal of this framework is to, in affect, standardize the kinds of 'design seeds' one might start with. Although any design is fair game for experimentation and evaluation, we require a certain kind of packaging to plant them.

One may be led to ask: “Is it enough to let anonymous users do usability testing on a system through the web, in order to call it a recycled research project?” The short answer is no, it is not. First, the conception that internet users are anonymous is faulty. We generally have a large amount of metadata about these kinds of users. Second, Recycled Research projects need to *actually be recyclable*. That is the components of the design need to be mechanically adjustable in a way that we can preserve a sense of meaning. It may be useful here to introduce an example of what we mean.

Recycled Canvas is the primary pilot for Recycled Research, testing the methodology at large before more flexible experimentation. It can be viewed as typical of what we expect a Recycled Research experiment to look like.

The experimental goal of the project is to perform a usability study on a number of different information presentation techniques by offering a customizable design for Wordpress, a popular blogging engine. The design offers customization options general to blogs, while retaining an overarching simplicity and generalizability in appearance. On install, users are able to select the placement of the sidebar, an area of the blog reserved for navigation, optional content, and widgets. They may choose a light or dark color scheme, customize the blog header image (or remove it) and choose one of four principle information views.

The number of experimental treatments in the structured experimental variables is seventy-two, a number that we would likely consider impractical in traditional lab settings. There are also some unstructured experimental conditions here. Each first-tier user will be generating their own content, of variable length and amount, with variable topics, style, and domain. We expect to see (and do see) large blogs, small blogs, blogs about technology, politics, and even crafts and bicycles. This mechanism allows us to test the design in a variety of contexts, helping to get a better holistic sense of the meaning of the design.

The experiment measures a number of usability goals such as: successful conversion of search terms, translation of visitors into feed subscribers, general 'stickiness' of the page - specifically amount of time spent and number of pages visited, the amount of article content read (or interpreted to be read) and so on. We also collect that actual mouse activity of the users - something that serves as a rough estimation of user goals, analogous to eyetracking. [Arroyo et al., 2006]

Users were given the opportunity to utilize an analytic tool for mouse tracking that allowed them to view user mouse tracks and metadata on their site (in the case of first-tier users) or other sites. AI processes are used to offer possible interpretation of the results, such as identifying mouse tracks indicative of reading. The discussion interface offered a basic forum style chat, with the ability to insert rendered analysis or data into the conversation with limited annotation support. We observed discussion to focus primarily on pattern identification of user behavior, bug or rendering errors, and comparisons between domain content. Users also noted the impact of customizations to the design outside the pre-seeded experimental options (such as color changes, etc.).

The mechanical requirements of recycled research formalize the capability to iterate on the design in such a way as to preserve the meaning of this evaluation discussion. This is only possible because of the continuous evaluation over time. When a change in the design occurs, new data can be compared to previous data. For example, turning the sidebar of the design red - a first-tier user design change - results in initial increases in mouse activity. Because of the open nature of the process, dissemination across new first-tier users of design changes allows the change to be explored in alternate contexts and built upon in different ways. Essentially we've constructed a design tree, where any design decision is a branch that has the possibility for growth. Human evaluation is responsible for creating design decisions, analogous with the necessary pruning of a tree.

Iterative Mechanical Growth

One quick observation here is that a lot of the design decisions one might do are fairly simplistic. Not every iteration is interesting. In a lot of design work there is a necessary amount of attention to detail, adjusting fonts, playing with colors and so on. A logical step in our process was adding the ability for the ‘trees’ to grow, at least in limited ways, without human intervention. This makes the metaphor of garden tending clearer. Humans are responsible for providing more than mechanical solutions and directing areas of growth. Even when the machine growth produces a truly novel design solution, recognition of its value by humans is responsible for its propagation.

What’s our basic primitive for web design, our single celled organism to evolve? One answer would be a web page. Web pages can be really complex. In fact, web pages can seem as complex as actual organisms we see around us. These organisms aren’t just giant cells. They’re composed of large numbers of small cells that reproduce and (occasionally) mutate on their own. This is more of what we’re looking for. What’s a single cell on a web page? It’s any html element.

We could essentially use every html element as a primitive. Every semantically meaningful element is just a specialized case of two genetic elements, a span or a div. A div describes a block level element and a span describes an inline element. In practice we can reduce this even further. A <span> is just a div with an inline style rule. That leaves us with a div as our single celled organism. A div can be modified by css to represent any element we can come up with on a normal web page. An img tag is just a div with a fixed height, width, and background image. Well, almost. Certain things are a little hard to do purely with css. We can't make a div an link, for example, or an iframe. We can mimic this behavior with a little bit of javascript or sever side code. Each of these divs has to exist in a shell of some sort - an environment. This environment becomes the rest of the web page.

DNA, Deoxyribonucleic acid, is the genetic instruction for the development of all (known) organisms. The genome is encoded in dna to allow the preservation of genes and traits (this is a vast oversimplification). For our divs, what describes them? There are only two properties of a div we might be concerned about. What's in it (the content) and what it is like (or what it looks like, the style). For us, css is dna, for all intents and purposes. Css describes a div very similar to the way our dna describes us. To end up with a different element, you have only to change its css. Elements with borders have the border property enabled. Elements with large text have a large font-size selection.

In our use of css as dna, we will break each property up as much as possible and limit them as much as appropriate. So we won't use just padding, we'll use padding-left, padding-right, padding-top, and padding-bottom. This will let us think of each one as an individual trait. Padding-left might be a useful mutation in an environment where no other divs have padding. It would set the element off visually from the others in a linear presentation. We may also impose caps on the range of possible values, such as setting the color property to only mutate to traditional web colors - otherwise there are a lot of possible values to worry about.

At this point we're pretty much going to ignore sexual reproduction. Why? The real problem with sexual reproduction for is it requires some mating preference relationship, more than one div and possibly some sense of proximity within a web page (divs should mate with nearby, similar, divs). It also would require a better understanding of the familiar hierarchy of a div. At some point two divs that have diverged enough should not be able to mate, the dna is just too different. This sounds like a lot of extra work which is not clearly supportive of our design goals, so we'll focus on asexual reproduction and mutation as our mechanism for evolving.

In our model a genetic div can reproduce, immediately die, and be replaced by its descendent offspring. It passes all of its genetic code onto its offspring. Basically the only work here that we need to do is decide how often reproduction occurs, the necessary requirements for reproduction, and what the chance for mutation is (otherwise we would never see any genetic change).

Death occurs when a div isn't fit enough to pass on its genetic material after its reproduction cycle has come up. At this point, we start over with a fresh seed div and the cycle of life begins again. We have a world (the web), so what is the necessary fitness for a div? Fitness implies that a div is good at living in the world - the genetic material is valuable and should be passed on. The easy answer is that divs live on mouse impressions - or more directly, divs live on attention. Attention is kind of hard to quantify. Optimistically we might say that attention implies a good, usable design. We will ignore the inner content of the div to some degree, since we have selected it to be both important and standard (in our test it will be the about sidebar text for a site).

These genetic divs allow us to explore design possibilities mechanically, as they mutate and iterate through low level design decisions.

A Design for Design Growth

We have expanded this to a general implementation procedure for the deployment of Recycled Research require that software applications conform to a model that follows six procedural stages:

1. Target Installation: The first stage involves the creation of software design as an easily embeddable or deployable instance.
2. Embedded (passive) information collection: Our goal is to measure existing human experience with the design. We rely on the kinds of implicit cues about human behavior that we can get through our limited medium.
3. Machine growth: The information collection provides the necessary fitness test for some level of pseudo-genetic iteration by the software itself through genetic divs, seeded by first-tier user design decisions. This expands the solution space for the design problem.
4. Local Data Collection, Harvesting and Aggregation: Human experiential accounts are collected locally, data is sampled, harvested and made available to the public.
5. Analytic Tools & Human Interpretation: The power of human computation is central to the value of this approach. After deployment, first-tier users are able to take advantage of the data and analytic tools to draw conclusions about the results. They can extend the work by copying or transplanting a genetic page, which in turn alters the machine growth pattern.

This creates a methodology focused on rapid design propagation with human mediated genetic modification to the design. This is possible largely because of the emphasis on a number of factors:

Viral Propagation - Our inquiry has been structured to be a decentralized model, with possibilities for viral spread and easy embedding.

Structuring Community Tending - Recycled research project goals (design goals, in this case) are set and adjusted by a community that observes the automated evolutionary iteration and makes high-level decisions based on associated data, analytic tools, and community discussion.

Organic Experimental Conditions - As the software is deployed virally, we require the host to set initial seed conditions. In a sense, this promotes organic or even market driven conditions. If a condition is not interesting or valuable to the community it will be used less, or even not at all.

Participatory Cost Model - The approach institutes a new kind of cost, a participatory one, requiring that any first-tier user of the software continue the experimental mechanism and allows iteration.

Genetic Iteration - While human analysis is necessary for the determination of high-level success, the iterative work can be combined with a genetic algorithm that explores alterations to the initial experimental conditions. Capabilities for genetic iterative adjustments provide growth direction.

This creates the basic structure for a garden of design solutions. The only remaining step is the necessary harvesting of promising ones.

A Question of Creativity

The question to evaluate our work is, does this kind of system actually produce better, more creative design solutions? Anecdotally it appears that this is in fact the case. We've explored an iteration of Recycled Canvas, called Organic Canvas, which does genetically iterate through solution paths. A lot of the machine growth solutions cover some of the common design iterations made by first-tiered users in the recycled canvas experiment. The solution paths explored, and the modifications made by first-tier experimenters tend to be significantly more complex. It is difficult to argue that they are better designs but they are more complex and less intuitive ones. Does the machine ever come up with truly novel designs when left to its own growth cycle? We noticed a handful of cases where that could be true.

The act of creation has changed [Koestler, 1969]. This is not a radical change. The nature of design is that it often lends itself to the incorporation of existing elements and approaches. Design is not a blank slate. The garden approach described here more clearly suggests that. The emphasis is even more clearly on the selection and manipulation of existing design templates. This not only provides more design inspiration it also allows offline design exploration. A designer can create a design and then return to it later to see how it has changed and how it has been evaluated by a large number of people. We feel that this is suggestive of the promise of machine iteration and pseudo-genetic algorithms for human tended design and creative acts. Conceptually we move from being the machinists of design to the stewards and gardeners of a machine guided growth process.

Related Work and Acknowledgements

We owe a large intellectual debt to work being conducted in human computation. Von Ahn has recently explored similar ideas with respect problems in artificial intelligence, extending on his work with Captchas to focus on the utilization of human intelligence to solve hard AI problems [2006]. Through the construction of games, such as Peekaboom, Verbosity, and Phetch, Von Ahn outlines some of the ideas that we incorporate in Recycled Research, particularly the idea of coupling web software with a research goal.

References

Arroyo, E., Selker, T., Wei, Willy. 2006. *Usability tool for analysis of web designs using mouse tracks*. In Proc. ACM Conference on Human Factors in Computing Systems.

Couper, M.P. 2000. *Web Surveys: A Review of Issues and Approaches*. The Public Opinion Quarterly.

Dorst, K., Cross, N. 2001. *Creativity in the design process: co-evolution of problem–solution*. Design Studies.

Doubleday, A., Ryan, M., Springett, M., Sutcliffe, A. 1997. *A comparison of usability techniques for evaluating design*. In Proc. ACM Conference on Human Factors in Computing Systems

Gibson, J. 1997. *The Theory of Affordances*. In Perceiving, Acting, and Knowing.

Goldberg, D., Holland, J. 1988. *Genetic Algorithms and Machine Learning*. Journal of Machine Learning.

Koestler A. 1996. *The Act of Creation*. London: Pan.

Von Ahn, L., Liu, R., Blum, M. 2006. *Peekaboom: A Game for Locating Objects in Images*. In Proc. ACM Conference on Human Factors in Computing Systems.