Breed Better Banners:
design automation through online interaction

Richard Gatarski
School of Business, Stockholm University

Preprint of an article accepted for publication in *Journal of Interactive Marketing 16*(1) published by John Wiley & Sons, Inc.

Abstract

The aim of this article is to explore the use of an automated design system based on a Genetic Algorithm to design banner advertising. Just as genes encode traits in living organisms and propagate the most successful traits by natural selection, so this algorithm propagates features of the most successful banner advertisements and extinguishes the least successful traits. In the current experiment at a major online store selling CDs, the algorithm interactively evolved 320 different Web-banners over a period of 23 days. While the algorithm exhibited some deficiencies, the results indicate that the method is viable. It created innovative designs that performed increasingly better than reference banners, improving the click-through rate from 0.68% for a standard banner to 3.1% by the 16th generation of design, without human intervention from the advertiser. The article concludes with a discussion of some conceptual issues and directions for future research.

Introduction

The advent of online media, such as the World Wide Web (Web), is a challenge to traditional methods for market communications and research. But the interactivity supported by the Web also opens up new opportunities, such as online research (Englis & Solomon, 2000). Regarding advertising on the Web, Novak and Hoffman (1997) have reviewed the practice and proposed a standardized terminology. Online advertising in the form of banners is by now an important tool and expected to grow in volume as well as money spent (Dahlén, Ekborn & Mörner, 2000; Hofacker & Murphy, 1998).

The basic problem addressed in this article is how to design a banner that maximizes the Click-Through Rate (CTR). CTR is typically defined as the number of user click-throughs divided by the number of times the banner has been exposed. See McLuhan (2000) for an example of the issue from the practitioners' point of view. Briggs and Stipp (1999) have researched different forms of online advertising. They found that such advertising, including banner advertising, is effective and that the creative execution plays a major role. Other research has focused on the effectiveness of banners (Briggs & Hollis, 1997), copy tests (Hofacker & Murphy, 1998) and factors that affect "Webad" visits (Raman & Leckenby, 1998). In these and similar studies the data have been manually analyzed with statistical instruments such as chi-square tests and multiple regressions. While the results are most interesting, the external validity may be negatively affected by data generated from small samples in laboratory-like experiments.

Chatterjee, Hoffman and Novak (1998) have proposed models that predict surfer behavior based on the analysis of vast volumes of "clickstream data" generated by log files on Web servers. They noted technical problems (imprecise log files, networked sites, visitor identification) that restrict the modeling scope. Hofacker and Murphy (2000) used such clickstream analysis to study how the number of banners in a site affected the CTR. Drèze and Zufryden (1998) studied the nature and magnitude of errors in Web-based advertising research caused by problems regarding the identification of unique visitors, cache memories and messages never read/received. They concluded that there were significant problems to be resolved before Internet advertising is “ready for prime time” and comparable with standard media. In sum to reliably judge the effectiveness of individual banners with validity amongst a myriad of page appearances is a formidable task.

Market researchers may apply multivariate data analysis to cope with complex setups. For almost 30 years conjoint analysis has been a popular approach for measuring consumer preference structures (Green & Srinivasan, 1990). Drawing from this knowledge Drèze and Zufryden (1997) proposed a Web-based conjoint methodology that eliminates biases of current advertising research methods and utilize possibilities of the new media. They also argued for methods that are unobtrusive, experimentally based, externally valid, based on large sample sizes, instantaneous and have low cost.

The aim of this article is to explore a different approach to some of the above problems, the use of a Genetic Algorithm to optimize banner design automatically. The idea underpinning this approach is that interactive technology is *not* standard media. In particular, the interactive tool discussed here allowed a banner to continuously and automatically re-design itself in *interaction* with its viewers.

The live experiment

One of the largest online music retailers in Europe, Boxman AB, participated as a site for the current experiment at its Swedish location boxman.se. The site handled close to 15,000 visitors every day. Before and during the experiment all visitors were randomly routed to one of five servers where they on average requested three pages. This routing is a common practice, as single Web servers have a limited visitor capacity. Boxman did not use any system (such as cookies) to keep track of individual visitors. Services from DoubleClick, Inc. managed both the banner space Boxman dynamically dedicated for in-site promotion and the banner space open for the DoubleClick advertising network.

A closer look at Boxman reveals a very complex structure. At the time boxman.se carried over 100,000 different products. Every page was dynamically created and contained navigation menus, search fields, product pictures, promotional text, top-10 lists, banners and more. This meant that a particular banner could be placed in one or more of millions of different pages. Later the Boxman system became even more complex as it managed nine different languages, currencies and country specific product lines. Thus it is here argued that the Boxman reality was far more complex than the environments researched in the above-referred studies. Boxman filed for bankruptcy in October 2000, over a year after the current experiment. That development does not affect the data, analysis and discussion presented here.

How to find the best banner design

The type of banners discussed here was used to promote individual products, a category of products, or campaigns. One of Boxman's art directors has developed a small set of "Boxman-styled" banner design concepts. The most common concept (468 x 60 pixels) used an illustration, two lines of animated text and a "click here" sign (see example in Figure 1). While tools for the Web make it easy to generate different designs, the literature above points to methodological problems in assessing their relative effectiveness.

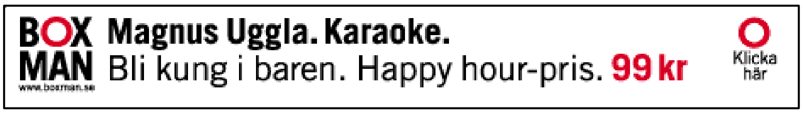

Figure 1. Example of a banner with the standard Boxman design style. Showed here is the last of 13 frames in an animation that gradually presents the banner contents and creates a blinking dot on the right. (The banner's color may not appear in print.)

The criterion used to assess effectiveness was CTR. While sales might have been a better criterion, Boxman's Web system was not able to relate banners to sales. The aim of the experiment was to compare interactively designed banners with banners designed in the normal Boxman style. This implies that the experiment should consider design objects atypical to

Boxman banners. Two measures were specifically taken to support external validity. To begin with, we choose the real artist *Robyn* and aligned our experiment timeframe with the real promotion of her latest album *My Truth*. Secondly, we wanted to stay with established norms for CD banner advertising. Therefore, we made a small survey of what kind of design objects such banners exhibited. These objects turned out to include pictures, graphics, artist names, album titles, price, endorsements, action verbs and different types of animation. We selected a range of objects that included illustrations (photographs or graphics) and copy texts with varying font characteristics. Due to limited resources we chose to stick with the standard layout for Boxman banners and varied only the objects, their placements and characteristics. For the same reason we had to limit the experiment to non-animated banners. In order to simplify the programming required to render the banner we divided the banner area into four parts as illustrated in Figure 2. Two parts formed as quadrants accommodated illustrations and two parts formed as rectangles accommodated copy texts.

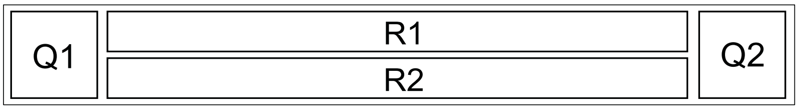

Figure 2. In order to simplify the programming the banner area was divided into two quadrants (Q1 and Q2) and two rectangles (R1 and R2). Illustrations (i.e. pictures or graphics) were placed in the quadrants and copy texts were placed in the rectangles.

Although these restrictions limited the modeling scope, a large number of different designs were still possible. A traditional conjoint analysis to test a set of 5 different photographs and 2 graphics in 16 different combinations plus 3 copy texts, each with 2 different fonts in 4 sizes and 2 colors, would have required 5x2x16x3x2x4x2=7,680 concepts. While a number of conjoint based methodologies exist to analyze setups with more than six attributes (Hensel-Börner & Sattler, 1999; Srinivasan & Park, 1997), they rely on multiple step approaches that demand longitudinal identification of the visitors, a requisite that Boxman could not fulfill.

Ideally an advertising research method should take into account ad clutter, message wear in and banner burnout (Chatterjee, Hoffman & Novak, 1998). Other important issues are the cost of implementation and the sophistication of the underlying methodology. High costs and complex methods might for example scare away practitioners (Carrol & Green, 1995). These effects and issues are considered in the next section and revisited further in the results section.

Multivariate analysis and Genetic Algorithms

Fortunately there are new techniques available for multivariate analysis (Hair, Anderson, Tatham & Black, 1998). Some are based on learning concepts found in Artificial Intelligence (AI) technologies. In a study of Direct Marketing data, Eiben et al. (1996) compared three such new methods. The authors found the methods equally accurate, interpretable and that they required the same amount of expertise and computer time. One of the methods, Genetic Algorithms (GAs), is inspired by the theory of evolution and survival of the fittest. This method was applied in the current experiment. For an excellent introduction to Genetic Algorithms see Mitchell (1998). Mühlenbein (1997) provides a more elaborated mathematical discussion of the issues within the field. Genetic Algorithms and other *Evolutionary Computation* techniques have been successfully used for many design problems (Lewis, 2001; Pollack et al., 2000; Soddu, 1999; Takagi, in press).

Within marketing the use of GAs is still limited. Hurley, Mouthino and Stephens (1995) outline 11 areas where GAs have been used in marketing management. However no previous study has specifically dealt with areas related to market communications. Furthermore the studies have only analyzed previously generated data. In contrast the current GA should create new data, analyze that data (learn about the results) and automatically improve the banner designs. The algorithm developed for the current experiment controls what design objects to use (including their characteristics) and where to place them in the banner.

Methodology

At this point it is useful to introduce some terminology associated with GAs. Designs are encoded as *chromosomes*, or collections of *genes* that build the blueprint for an *organism* (in this case the banner). Each gene encodes a specific *trait* (for example the amount of copy text). The various possible settings for a trait are called *alleles* (for example 12, 14, 16 or 18 point type).

The rest of this section provides a detailed description of the chromosome, important notions regarding the programming of the algorithm and lessons learned from the implementation. Besides a limited time frame and a low budget we had little knowledge about genetic algorithms and how they could be used in live Web-interaction. Furthermore we were dealing with a commercial system running at full speed. This restricted us from making some useful revisions to the chromosome as well as the algorithm.

It is important to note that the subsequent description is based on the actual outcome of our programming efforts, which deviates somewhat from what we originally intended. As a result the technical particulars appear rather cluttered. Even so, for analytical purposes as well as for future research, we believe it is useful to provide this level of detail. Important consequences from our setup are further elaborated in the limitations section at the end of the article.

Description of the chromosome and the algorithm

A growing number of software packages that deploy genetic algorithms are publicly available. To our knowledge none of these could easily be used for our purpose. Therefore we decided to program our algorithm from scratch and base it on a single bit-string chromosome. While different ways to encode the solutions are possible (e.g. many-character, real-value and tree encodings), fixed length and fixed ordered bit strings are often used (Mitchell, 1998, pp. 156-158). One single bit can only encode two different settings, therefore traits with more settings require more bits, and thus our chromosome was made up of a string of all necessary bits. Traits for all feasible bit combinations for each gene were defined in order to eliminate impossible concepts. Furthermore we sought to eliminate ugly combinations (e.g. the price before the artists name or different font sizes within one rectangle). The 51-bit chromosome, with its 17 genes, is schematically described in Figure 3.

In the current experiment only 40 bits were active, that is, contributed to variations in the banner design. The IC gene controlled how the illustrations can be combined (2 bits). The CC gene did the same for the copy texts (4 bits). The IA and CA genes controlled how the illustrations and the copy texts were arranged on the banner (1+1 bits). The AP gene controlled which artist picture to use (2 bits). The AV gene controlled which action verb to use, its font face, size and color (6 bits). The ED did the same for the endorsement (6 bits). The artist's name, title and price were fixed, but the AN/TI/PR genes still controlled their appearance (4+4+4 bits). The PT gene controlled which price type to use, its font face and color (4 bits). Genes G1 and G2 controlled which graphic to use (1+1 bits). While the last two genes had identical alleles, the IC gene used them independently.

The inactive bits are explained as follows. At the outset we planned to also vary banner background color (1 bit), artist name (2 bits), product description (1 bit), title of the album (2 bits), product picture (1 bit) and price (2 bits). That ambition was reduced, but not until after the chromosome was designed. Furthermore, as a result of a coding error bit 39 was redundant (1 bit). To determine the exact effect of this error would require considerable efforts. Possible effects are that only one of the available typefaces were possible; that the system default typeface was used (no indication of that in the rendered banners); that the distribution of typeface choice was 25/75; or that the distribution was 50/50, that is the error did not matter. We believe that this error has no significant impact on the experiment as discussed here. Finally, bit 50 was allocated, but never used (1 bit).

The algorithm started by randomly generating (flipping bits) within an initial population of 20 chromosomes. Banners were then rendered using information in the chromosomes and were exposed to the *environment* (visitors at the site). At a certain point in time, which is discussed subsequently, some chromosomes were selected for reproduction based on their *fitness*, which in this case meant how well the chromosome solved the problem of attracting click-throughs. In this algorithm the fitness (probability for a banner to be selected as a parent) was directly proportional to its CTR. Using this criteria two parent chromosomes were drawn from the current generation. If the two chromosomes were the same, two new ones were drawn.

Bit pos. and size	Gene	Banner trait	Allele
0: 1	BC	Background color (inactive)	White, White
1: 2	IC	Illustrations Combination	AP│PP, AP│G1, PP│G1, G1│G2
3: 4	CC	Copy Combination	AN/TI, AN/ED, AN/AV, AN/PR, AN/PT&PR, TI/ED, TI/AV, TI/PR, TI/PT&PR, ED/AV, ED/PR, ED/PT&PR, AV/PR, AV/PT&PR, PT&PR/-, PR/-
7: 1	IA	Illustrations Arrangement	No Swap, Swap left/right quadrant
8: 1	CA	Copy Arrangement	No Swap, Swap upper/lower rectangles
9: 2	AP	Artist Picture	
11: 1	PP	Product Picture (inactive)	
12: 6	AV	Action Verb (copy:2, face:1*, size:2,:color:1)	Click to play Play, Hear her Truth Main Thing Recovered, Köp nu!
18: 6	ED	Endorsement (copy:2, face:1, size:2,:color:1)	Stevie Wonder Dig, Top 10 placering Bland det vackraste, Storstadshit
24: 6	AN	Artist Name (copy:2, face:1, size:2,:color:1) (2 copy bits inactive)	Robyn, Robyn, Robyn, Robyn
30: 6	TI	Title (copy:2, face:1, size:2,:color:1) (2 copy bits inactive)	My Truth, My Truth, My Truth, My Truth
36: 5	PT	Price Type** (copy:2, face:2,:color:1) (1 face bit inactive)	Electric price, Urban price, Sant pris, Rytmiskt pris
41: 6	PR	Price (copy:2, face:1, size:2, color:1) (2 copy bits inactive)	129 kr, 129 kr, 129 kr, 129 kr
47: 1	PD	Product Description (inactive)	n/a
48: 1	G1	Graphic 1	Klicka här, BOX MAN
49: 1	G2	Graphic 2	Klicka här, BOX MAN
50: 1	SP	Spare (inactive)	n/a

*Figure 3. Description of how the banner traits were encoded into genes and the alleles available in the current experiment. The chromosome consists of 51 bits, of which 40 were active. [*For all copy texts typeface was either Trade Gothic Bold or Trade Gothic Bold Condensed No. 20 from Adobe. Font size was 12, 14, 16 or 18 points. Font color was either black (#000000) or red (#cc0033). ** Size for PT follows size for PR and due to a programming error the PT typeface gene used 2 instead of 1 bit.]. (The illustrations' color may not appear in print.)*

At this time a *crossover* point (between any two bits in a chromosome) was randomly determined. A child was created by combining the bits to the left of the crossover point from one of the parents with bits to the right of the crossover point from the other parent. In addition the child chromosome was *mutated* by randomly selecting five bits in the chromosome, each having a 50% probability of being flipped. This process, which included returning the parents to the drawing pot, was repeated until a new generation of chromosomes had been generated (20 new chromosomes). The newborn banners were then exposed on the site and the algorithm continued until it was manually stopped.

Banner number	*Exposures*	*Clicks-throughs*	*CTR (%)*
7	1073	15	1.4
17	1073	15	1.4
6	1073	14	1.3
3	1073	13	1.2
8	1073	13	1.2
11	1073	12	1.1
1	1073	11	1.0
18	1073	11	1.0
20	1073	11	1.0
16	1073	10	0.9
19	1073	10	0.9
2	1073	9	0.8
12	1073	8	0.7
5	1073	8	0.7
15	1073	8	0.7
14	1073	8	0.7
4	1073	7	0.7
13	1073	7	0.7
10	1073	7	0.7
9	1073	7	0.7

Table 1. Click-Through Rates (CTR) for the first generation of Evolving Banners. The table is sorted by CTR in descending order

A Web-programmer developed and implemented the algorithm in C++. The chromosomes were automatically rendered to GIF-banners with a standard software application. We carefully checked that the algorithm at least seemingly bred banners based on earlier generations. Through appropriate HTML-encoding we made sure that new material replaced banners stored in browser cache memories as well as proxy servers. Table 1 shows data generated by the algorithm for the initial population of banners. In this first generation for example banner number 7 was requested 1073 times and received 15 click-throughs. Figure 4 and Figure 6 illustrate banners designed by the algorithm.

Figure 4. Two examples of banners from the 8th generation in the Evolving Banners category. They illustrate the "design themes" that emerged in this generation. (The banner's color may not appear in print.)

Control banners and expected results

Boxman had organized their pages into different types (CD pages, DVD pages, new arrivals pages, search result pages, etc.). For example, when a visitor searched for an artist the search results page contained a banner in addition to information about found products. Boxman generously allocated 40% of their banner space at about 10 such page types. We divided that space into three categories. The first category was used for the evolving populations (20 banners in each generation). The other categories were used for control purposes. The second category was used for an exact replica of the first banner population (20 randomly designed, but not evolving). Finally, the third category was used for a banner that was manually designed according to the Boxman standard style, except that it was not animated. These categories are henceforth referred to as Evolving Banners, First Banners and Standard Banner.

Every time the site got a request to submit a banner within the allocated page space, 40% of the requests were randomly distributed among the categories as illustrated in Figure 5. The remaining 60% were forwarded to the DoubleClick advertising system. Thus the current experiment was comparable to a live experiment with an unobtrusive measure of a census population (page requesters).

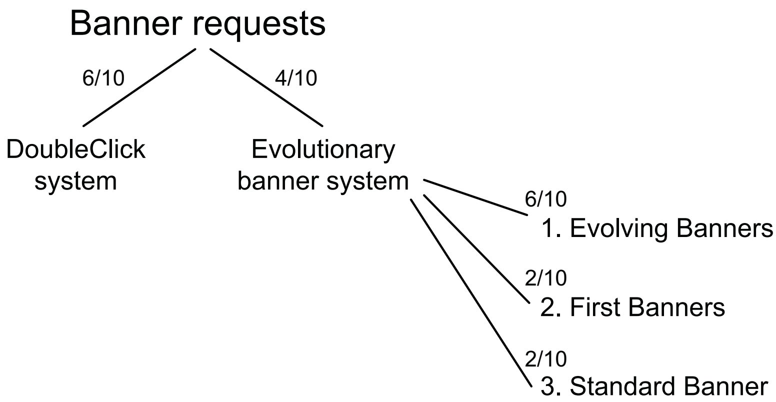

Figure 5. How the server routed requests for banners. Four out of ten requests were directed to the experiment system. Remaining requests were forwarded to the DoubleClick network.

A few expectations based on initial test runs with other products were outlined before the current experiment. The first expectation was that the Standard Banner (manually designed) would outperform First Banners (randomly designed). The second expectation was that Standard Banner would initially outperform the Evolving Banners, but that would change as the latter evolved. Main limitations with GAs are that they require large populations and time (many generations and exposures) to evolve (Hair et al., 1998; Mitchell, 1998). Therefore the third expectation was that this could be realistically implemented on the Web, here within boxman.se.

Results

The limitations of the current execution are such that the findings in this study need to be interpreted with caution, and a few details must be noted before the current data can be analyzed. During the experiment Boxman servers had multiple downtime periods, producing more click-throughs than specified on four occasions and forcing us to re-start the algorithm a number of times. This might have caused wear-out effects for banners that were exposed before the experiment begun. In addition, the results are tentative as the banners probably needed more exposure (explained further below) and the algorithm exhibit some deficiencies. Therefore the presentation of our results is focused more on conceptual issues than empirical outcomes.

Campaigns within boxman.se that promoted CDs with the use of standard design banners typically ran for about two weeks. On average that gave around 25,000 exposures per banner, with a Click-Through Rate (CTR) of slightly less than 1%. The current algorithm created 16 banner generations before it was stopped after 23 days running time in mid August 1999. During that period the banners received over 330,000 exposures. Their distribution is specified in Table 2. The Standard Banner had a fairly stable CTR of around 0.68% from the first day. That is slightly less than the boxman.se average, possibly because the banner was not animated and our system did not filter out request from non-humans (e.g. spiders and robots). However it indicates that the CTR data are normal, suggesting that the measuring system worked.

At the end of the experiment, the randomly designed First Banners had an average CTR of 1.00%, well above the 0.68% for the Standard Banner. This was quite the opposite of the first expectation. Furthermore the best performing banner among the First Banners had a CTR of 1.70% (data not shown in table), and the best 5 averaged 1.50%. Three possible explanations for this relatively better performance are suggested. One, the randomly designed banners that perform well attracts more click-throughs because they are relatively “unbranded” as argued by Briggs and Hollis (1997). Two, distinctive designs do by themselves attract visitors/preferences and thus more click-throughs. Three, the better performing banners had designs that are more geared towards click-throughs than the standard design.

An analysis of First Banners over time (data not presented here) revealed another unexpected finding. Initially (after about 100 exposures per banner) the individual banner CTR varied between 0% and 4.7%. Later (after about 1,000 exposures per banner) that variation was reduced to between 0.3% and 2.3%. After that point the First Banners average CTR always outperformed the Standard Banner. When the experiment was stopped, the CTR variation among the First Banners was further reduced to between 0.6% and 1.7% (about 3,400 exposures per banner). This suggests that CTR data are unreliable before approximately 1,000 exposures per banner.

Category	*Generation*	*Exposures*	*Click-throughs*	*Average CTR all banners (%)*	*Average CTR best 5 banners (%)*	*"Click here" occurrence*
Evolving Banners	1	21465	204	0.95	1.3	9
	2	14660	150	1.02	1.5	14
	3	13500	173	1.28	1.9	12
	4	10460	150	1.43	2.0	16
	5	11622	150	1.29	2.3	16
	6	11820	150	1.27	2.1	18
	7	14820	150	1.01	1.5	16
	8	14160	150	1.06	1.6	16
	9	14480	204	1.41	2.6	18
	10	10280	150	1.46	2.2	18
	11	9600	150	1.56	2.6	21
	12	10200	150	1.47	2.4	20
	13	10120	167	1.65	3.2	20
	14	11460	151	1.32	2.3	21
	15	9635	150	1.56	[a]	22
	16	9040	150	1.66	3.1	20
First Banners		68780	691	1.00	1.50	9
Standard Banner		69138	473	0.68		1

Table 2. Summary of exposures and click-throughs for all banners during the experiment. Exposure shows how many times all banners in the category/generation were requested. Click-through shows the sum of all click-throughs for the banners in the generation/category. "Click here" occurrence shows the number of "click here" signs found in all banners in the generation/category. [a]Individual banner data not available.

The second expectation was that Evolving Banners would outperform the Standard Banner. The data in Table 2 shows that the average CTR improves from generations 1 to 4, then degrades for the next 3 generations, and finally improvements return. As can be seen the average CTR of the best 5 performing banners in each generation is around twice as high as for the control banners. It would therefore be tempting to settle with the 5 best performing banners in each generation. But the algorithm depends on all banners. Individuals who perform badly might yet carry valuable traits. In other words, it is the population average that must be considered. The numbers indicate that the method is effective, that is, creates increasingly better-performing banners.

We now discuss three arguments for further experiments with this method. The first argument is based on a visual inspection of the banners produced by the algorithm. The second argument follows the fact that the algorithm needs fine-tuning. The third argument concerns burnout effects and the nature of GAs.

Visual Inspection

A visual inspection of the first generation of banners (leftmost column in Figure 6) gives the impression of random design variation. When we looked at the eight generation, one thing was striking - the designs had converged into five themes. Figure 4 illustrates two of those themes, both occur with slight variation in five banners each. Hypothetically this would mean that the algorithm was converging on a few design types. The way certain objects were favored also caught our attention. For example the number of "click here signs" are summarized in Table 2. That number quickly went up from 9 to 18 and then stayed above that level. One possible interpretation is that the algorithm learned what we already knew - academic research as well as practitioner experience is that the expression "click here" on a banner often increases the CTR. To discuss why different banners perform differently is beyond the scope of this article. Still it is most interesting to compare the designs in the first and last generation of banners (see Figure 6). As can be seen the algorithm has eliminated all pictures except those with a close-up on the artists face. The copy text "Click to play Play" occurs 10 times in the final generation, even though it was not present in the first generation. As a matter of fact it showed up in the fifth generation. We leave it to readers to judge whether or not the evolved designs are in harmony with established knowledge about what gets attention and click-throughs.

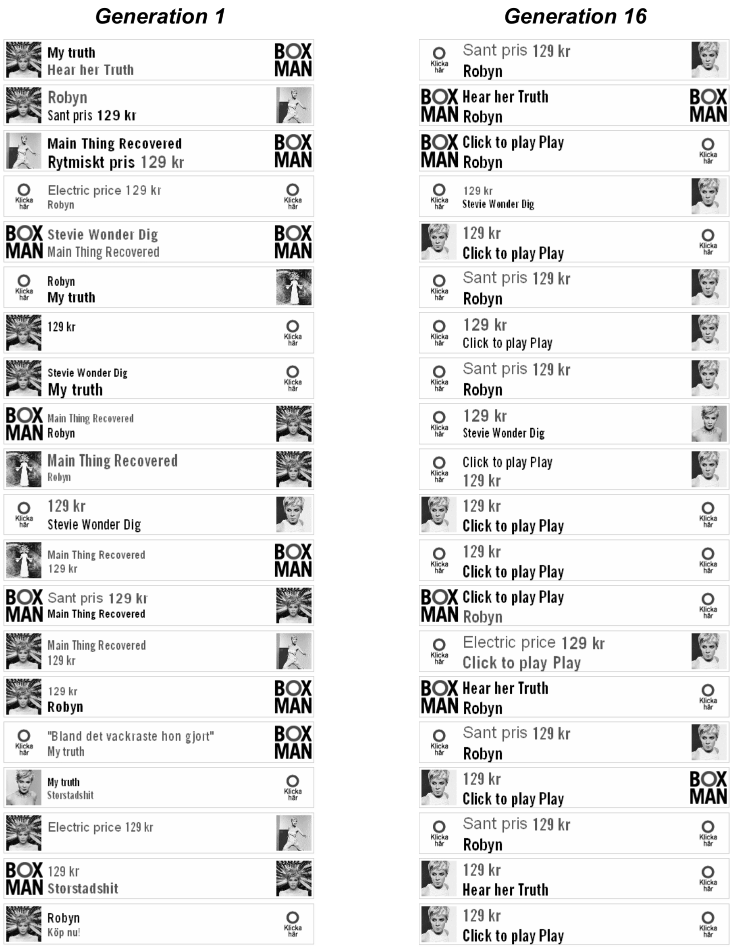

Figure 6. All banners from the first and last generations in the Evolving Banners category. When displayed at the site, each banner is 468 x 60 pixels. (The banners' color may not appear in print.)

Fine-tuning of the algorithm

How frequently should new generations be created? We considered five rules. First, *Time*: After a certain time, for example 24 hours. Second, *Population click-throughs*: After a specific number of click-throughs on the current generation, for example 150. Third, *Best banner*: After any banner from any group has received a specified number of click-throughs, for example 50. Fourth, *Exposure*: After every banner has been exposed a specific number of times, for example 1,000. Fifth, *Reliability*: The algorithm uses an appropriate method to judge if the data are reliable enough. Based on the test runs and a rough estimate of the available banner space we decided to apply the second method. Thus fitness was computed when the sum of click-throughs in a population reached 150.

Data in Table 2 shows that every banner in the first generation aggregated 1,073 exposures. This is close to the requirement according to the reliability discussion above, but caused by a server error which ran the population beyond 150 click-trough's. When that error was fixed and as the banners improved, they required less exposure to reach the breeding point. This suggests that we should have used another rule, preferably the fifth, which is based on reliability. After a second analysis of First Banner performance over time we believe that the algorithm should have based the fitness on data with bad, but not terrible, reliability.

One way to give each banner more exposure is to reduce the population size. But that would result in a smaller genetic variance. Another way is to wait for more click-throughs, which takes time. It has been suggested that a GA should typically run for anywhere from 50 to 500 or more generations (Mitchell, 1998). This experiment was stopped after only 16 generations. Academic researchers could of course allocate more time and wait for better results. But, for practitioners the evolution rate is more critical. Products might be discontinued before the evolution catches up. For example, Boxman sometimes used the previous day's sales statistics to control which product to promote. In such cases there was no time for evolution. In sum, if the evolution rate is too low, the banner might not be able to evolve into an effective design.

Burn-out

The expression "banner burn-out" is frequently used amongst practitioners to denote the wear-out effect caused by repeated exposures of the same banner to consumers (Chatterjee, et al. 1998). At least in theory, the method applied here would address this problem in a “natural way”. For example, the algorithm would automatically replace a banner when it burns out, that is, its CTR decreases. To this can be added that the continuous redesign of banners might also postpone or reduce the wear-out effect.

Conclusions, limitations, and future research

The aim of the current experiment was to explore new possibilities for Web advertising design such as those proposed by Drèze and Zufryden (1997), as well as address many of the methodological problems reviewed above. This article proposed the use of new methods based on evolutionary computation to improve the effectiveness of online advertising. Such a method was also applied in the current experiment.

It was found possible to implement a Genetic Algorithm, here with a population of 20 banners and 23 days time to evolve. Tentative results indicate that the method was effective. The system did create innovative designs that performed increasingly better than control banners. The methodology is easy to understand and was implemented with relatively low effort. These were the first steps in a new direction with little established knowledge, such as rules of thumbs.

Due to the exploratory character of our approach, some deficiencies with the used algorithm may have influenced the data quality in a negative fashion. Unfortunately, because of budget limitations and the fact that Boxman AB ceased their operations, it was impossible to attend to these issues and collect new data. Besides what have been noted above, two deficiencies in particular need to be clarified here.

First, the oversized and messy character of the chromosome implies an unnecessary large universe of solutions, as well as it makes it difficult to understand exactly how the chromosome is reproduced. The algorithm rendered nowhere near all of the possible concepts (designs). The concepts used were only the ones generated in the first randomly designed population together with those that resulted from random crossover and mutation. A dependence on unexplored randomness is probably a disadvantage with this method. Still, established statistical methods are often based on different acknowledgments of random variation. Future research should address this limitation, and thus compare and advance the question of randomness when applied on interactive design processes similar to the current experiment.

Second, due to a misunderstanding between the designer of the experiment and the programmer, the mutation rate is much greater than intended. Mühlenbein (1997) has in many experiments examined the effects of various GA setups, including recombinations (crossover) and mutation rates. His findings suggest that our high rate might reduce, rather than destroy, the effectiveness of our algorithm.

In addition, the method used in the current experiment is geared for mass advertising, not Customization (Pine, 1993) or One-To-One marketing (Pepper & Rogers, 1993), which are important themes in contemporary marketing. Still mass communication is required in many situations, such as when the visitors are anonymous, sites attract new visitors, the product is based on a broad audience (news, music) or customization is not feasible.

The speed with which new banner populations is generated, the *evolution rate*, is crucial in a fast and dynamic environment. Appropriate ways for fine-tuning of the algorithm constitute an interesting domain for future research. This includes chromosome design, population size, fitness calculation, crossover parameters, mutation rate, etc. Many of these could even be evolved with GAs (Mitchell, 1998, p. 174).

Banners are typically used to attract visitors to a site, or to promote offerings within a site. In the former case mass advertising banner campaigns may have millions of exposures per day. In the latter case, which was considered in the current research, the number of exposures is generally much smaller. This kind of problem imposed some restrictions on how the algorithm could be implemented. To explore the possibilities with less limited restrictions, enabled by larger volumes of exposures, is an important subject for future research.

While the banner is one form of market communication, digital media employ other forms of communication that need design. We believe that many of those forms could be automatically and interactively designed. Future research could experiment with GAs in order to design for example interstitials, copy text in e-mails and campaign sites. Even products could be interactively designed with the methodology used here. Examples include physical goods in the concept stage, service packages and digital products such as news, music and computer software.

The current experiment used explicit interaction, here click-through. Other possible approaches are implicit interaction such as time spent or products sold. Further research is needed to explore these and other types of interaction that controls the fitness evaluation.

Finally, scientists usually aim for understanding, explanation and prediction. We want results and the ability to generalize those results. Models based on Artificial Intelligence technology, including GAs, are on the other hand typically good for exploration and application. Hofacker and Murphy remark on how easily large volumes of data can be generated and analyzed in the Web environment (2000). But they also acknowledge the difficulties that arise when results are generalized to the universe of possible pages.

It is important to note that the result from this experiment is *not* the particular banner designs. Instead the result is a particular algorithm used to automatically and interactively design banners. Hopefully this algorithm could be generalized to other online advertising forms, products and environments. What happens is that we no longer ask *why* a particular banner design works, but *how* the algorithm could be improved. These consequences have to be discussed in order to assess the scientific progress with new methodologies based on automated and interactive design.

References

Briggs, R. & Hollis, N. (1997). Advertising on the Web: Is there response before click-through? *Journal of Advertising Research, 37*(2), 33-45.

Briggs, R. & Stipp, H. (1999). How Internet advertising works. In *The worldwide Internet conference Net Effects (Vol. 227)*. London: ESOMAR Publication Series. Retrieved August 07, 2001, from http://www.mbinteractive.com/word_downloads/esomar.doc

Chatterjee, P., Hoffman, D. L. & Novak, T. P. (1998). *Modeling the Clickstream: Implications for Web-Based Advertising Efforts*. Retrieved October 05, 2000 from Vanderbilt University, the eLab Web site: http://ecommerce.vanderbilt.edu/papers/pdf/clickstream.pdf

Carrol, J. D. & Green, P. E. (1995). Psychometric methods in marketing research: part I, conjoint analysis, *Journal of Marketing Research, 32*, 385-391.

Dahlén, M., Ekborn, Y. & Mörner, N. (2000). To click or not to click: an empirical study of response to banner ads for high and low involvement products. *Consumption, Markets & Culture, 4*(1), 57-76.

Drèze, X. & Zufryden, F. (1997). Testing Web Site Design and Promotional Content. *Journal of Advertising Research, 37* (2), 77-91. Retrieved July 5, 1999 from http://sbaxdm.usc.edu/Publications/jar.pdf

Drèze, X. & Zufryden, F. (1998). Is Internet Advertising Ready for Prime Time? *Journal of Advertising Research*, 38(3) Retrieved July 5, 1999 from http://sbaxdm.usc.edu/Publications/webgrp.pdf

Eiben, A. E., Euverman, T. J., Kowalczyk, W., Peelen, E. Slisser, F. & Wesseling, J. A. M. (1996). Comparing Adaptive and Traditional Techniques for Direct Marketing. In H. J. Zimmermann (Ed.), *Proceedings of the 4th European Congress on Intelligent Techniques and Soft Computing* (pp. 434-437). Aachen, Germany: Verlag Mainz. Retrieved August 7, 2001 from http://www.wi.leidenuniv.nl/~gusz/eufit96.ps.gz

Englis, B. G. & Solomon, M. R. (2000). Life/Style OnLine(c) a web-based methodology for visually-oriented consumer research. *Journal of Interactive Marketing, 14*(1), 2-14.

Green, P. E. & Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *Journal of Marketing, 54*(4), 3-19.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W.C. (1998). *Multivariate Data Analysis*. New Jersey: Prentice Hall.

Hensel-Börner, S. & Sattler, H. (1999). Validity of the customized computerized onjoint Analysis (CCC). In *Proceedings of the 28th EMAC conference in Berlin*. Retrieved from CD-ROM published by Humboldt University, Faculty of economics and business administration.

Hofacker, C. F. & Murphy, J. (1998). World wide web banner advertisement copy testing. *European Journal of Marketing, 32*, 703-712.

Hofacker, C. F. & Murphy, J. (2000). Clickable world wide web banner ads and content sites. *Journal of Interactive Marketing, 14*(1), 49-59.

Hurley, S., Mouthino, L. & Stephens, N. M. (1995). Solving marketing optimization problems using genetic algorithms. *European Journal of Marketing, 29*(4), 39-56.

Lewis, M. (2001). *Visual Aesthetic Evolutionary Design Links*. Retrieved August 7, 2001 from Ohio State University, personal Web page: http://www.cgrg.ohio-state.edu/%7Emlewis/aed.html

McLuhan, R. (2000, June 22). Ways to make the clicks measure up, *Marketing*, 35.

Mühlenbein, H. (1997). Genetic Algorithms. In A. Aarts & L. Lenstra (Eds.), *Local Search in Combinatorial Optimization* (pp. 131-171). Chichester: John Wiley & Sons. Retrieved July 3, 2001 from ftp://borneo.gmd.de/pub/as/ga/gmd_as_ga-96_03.ps

Mitchell, M. (1998). *An introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Novak, T. P. & Hoffman, D. L. (1997). New Metrics for New Media: Toward the development of Web Measurement Standards. *World Wide Web Journal, 2*, 213-246. Retrieved October 5, 2000 from http://ecommerce.vanderbilt.edu/novak/web.standards/webstand.html

Pepper, D. & Rogers, M. (1993). *The One To One Future: Building Relationships One Customer at a Time*. New York: Currency/Doubleday.

Pine II, J. B. (1993). *Mass Customization: The New Frontier in Business Competition*. Boston, MA: Harvard Business School Press.

Pollack, J. B., Lipson, H., Ficici, S. Funes, P., Hornby, G. & Watson, R. A. (2000). Evolutionary Techniques in Physical Robotics. In J. Miller (Ed.), *Evolvable Systems: from biology to hardware; proceedings of the third international conference (ICES 2000) (Lecture Notes in Computer Science; Vol. 1801)* (pp. 175-186). Springer. Retrieved September 6, 2000 from http://www.demo.cs.brandeis.edu/papers/ices00.pdf

Raman, N. V. & Leckenby, J. D. (1998). Factors affecting consumers "Webad" visits. *European Journal of Marketing, 32*, 737-748.

Soddu, C. (Ed.). (1999). *Proceedings of the International Conference in Generative Art´98*, Rome, Italy: Dedalo Publisher.

Srinivasan, P. & Park, C. (1997). Suprising robustness of the self-explicated approach to customer preference structure measurement. *Journal of Marketing Research, 34*. 286-291.

Takagi, H. T. (in press). Interactive Evolutionary Computing: fusion of the capabilities of EC optimization and human evaluation", *Proceedings of the IEEE*. Retrieved March 20, 2001 from http://www.kyushu-id.ac.jp/~takagi/TAKAGI/IECpaper/ProcIEEE_3.pdf