Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy

Steven P. Dow[1], Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz[2], Scott R. Klemmer[1]
Department of Computer Science[1] and School of Education[2]
Stanford University, Stanford, CA 94305
[spdow, alana11, jkass, schwarz1, danls, srk]@stanford.edu

ABSTRACT
Iteration can help people improve ideas. It can also give rise to fixation — continuously refining one option without considering others. Does creating and receiving feedback on multiple prototypes in parallel — as opposed to serially — affect learning, self-efficacy, and design exploration? An experiment manipulated whether independent novice designers created graphic Web advertisements in parallel or in series. Serial participants received descriptive critique directly after each prototype. Parallel participants created multiple prototypes before receiving feedback. As measured by click-through data and expert ratings, ads created in the Parallel condition significantly outperformed those from the Serial condition. Moreover, independent raters found Parallel prototypes to be more diverse. Parallel participants also reported a larger increase in task-specific self-confidence. This paper outlines a theoretical foundation for why parallel prototyping produces better design results and discusses the implications for design education.

Author Keywords
Prototyping, iteration, feedback, juxtaposition, comparison, design, divergence, exploration, critique, self-efficacy

ACM Classification Keywords
H.1.m. [Information Systems]: Models and Principles

General Terms
Experimentation, Design

INTRODUCTION
Iteration is central to learning and motivation in design [20,32,51,52]. Yet, its primary virtue — incremental, situated feedback — can also blind designers to other alternatives, steering them to local, rather than global, optima [12,20]. To combat this, creating multiple alternatives in parallel may encourage people to more effectively discover unseen constraints and opportunities [16], enumerate more diverse solutions [12], and obtain more authentic and diverse feedback from potential users [58]. While a parallel approach has potential benefits, it can take time away from refinement.

Effectively educating a more creative workforce requires understanding how and why design practices affect results. Towards that goal, this paper investigates the relative merits of parallel and serial prototyping under time constraints. In a between-subjects experiment, thirty-three participants designed Web banner advertisements for a magazine. In both conditions, participants created five prototype ads and then a final ad. They received descriptive critique on each prototype. Participants worked independently and were given equal time to create each prototype and read each critique; the *structure* of the process differed across conditions. In the Serial condition, participants received feedback after creating each prototype. Participants in the Parallel condition created three prototypes, received feedback on all three, then made two more prototypes, and received feedback again before creating a final ad design (see Figure 1).

Figure 1 The experiment manipulates when participants receive feedback during a design process: in Serial after each design (top) versus in Parallel on three, then two (bottom).

The study measured design performance by running a MySpace.com advertising campaign with all participants' final ad creations and measuring click-through analytics. Independent experts also rated ad quality. To measure the diversity of each participant's ad creations, independent online raters judged pair-wise similarity between each of the participants' six ad prototypes. A self-report assessment measured participants' pre- and post-task view of task-

Author's manuscript: To appear in ACM Transactions on Human-Computer Interaction 17(4), 2010.

specific self-efficacy [24,31] (see Appendix B). The study concluded with an open-ended interview (see Appendix C).

Parallel participants outperformed Serial participants by all performance measures: click-through rates, time spent on the target client website, and ratings by the clients and ad professionals. Further, independent raters found that the diversity of each participant's prototypes was greater in the Parallel condition. Parallel participants reported a significant gain in self-efficacy, a measure of task-oriented confidence. Serial participants did not. In post-task interviews, nearly half of serial participants reported negative reactions to critique of their prototypes; no Parallel participants reported this. About half the participants had prior graphic or ad design experience. Participants with prior experience outperformed novices.

The study found that a parallel prototyping approach yields better results, more divergent ideas, and that parallel prototypers react more positively to critique. The results could significantly impact both how people approach creative problems and how educators teach design.

THEORETICAL BENEFITS OF PARALLEL DESIGN

Research on human problem solving traditionally examines problems with an optimal solution and a single path to reach that solution [44]. In design, problems and solutions co-evolve [19], constraints are often negotiable [51], subproblems are interconnected [28], and solutions are not right or wrong, only better or worse [48]. How and when to explore or refine solutions to open-ended problems remains an active debate in design research and education [7,16,45]. Without exploration, designers may choose a design concept too early and fail to identify a valuable direction [15]. Without refinement, ideas may not reach their full potential [8]. Navigating a design space may come easier as designers develop intuition, however even experts can exhibit fixation [15] and groupthink behaviors [34]. The architect Laseau posits an idealized conceptual model for exploring and refining, where designers iteratively diverge and converge on ideas, eventually narrowing to a best-fit concept [41]. This paper investigates the hypothesis that parallel prototyping increases learning, exploration, and design task confidence. More broadly, this research seeks a richer theoretical understanding of creative work to help practitioners and students design more effectively.

Parallel Prototyping Promotes Comparison

Throughout life, people learn interactively, trying different actions and observing their effect in the world [29,47]. Life provides a corpus of experiences from which to draw comparisons in new learning situations [38,54]. Examples can aid problem solving [4,33,42,55], especially when people explicitly extract principles [27,57]. Comparison helps people focus on key relations [26], aiding the acquisition of underlying principles [14,25] and sharpening categorical boundaries [11]. This paper hypothesizes that parallel prototyping better enables people to compare feedback on multiple prototypes, leading to a better understanding of how key variables interrelate.

Hypothesis 1: *Parallel prototyping leads to feedback comparison and produces higher quality designs*.

In the ad design study, quality is measured with click-through analytics and expert ratings.

Parallel Prototyping Encourages Exploration

The open-ended nature of design problems often requires designers to imagine and try out alternative solutions [12,36]. Without sufficient exploration, design teams may fixate on potential solutions [21,35], overlook key insights [37], make poor choices to justify prior investments in money or time [6], and exhibit *groupthink*, a "deterioration of mental efficiency, reality testing, and moral judgment that results from in-group pressures" [34]. Numerous interventions have been proposed to help designers think divergently, laterally, or "outside the box" [10,23,59].

Osborn posited *premature evaluation* as a major block to organizational creativity and proposed "rules" for brainstorming: think broadly early on and save critique for later [46]. Immediate feedback sets the focus on refinement, whereas postponing critique until after creating multiple designs encourages more divergence.

Hypothesis 2: *Parallel prototyping results in more divergent concepts*.

In the ad design study, independent raters judge the diversity/similarity of participants' sets of prototypes, providing a measure of design divergence.

Parallel Prototyping Fosters Design Confidence

Self-efficacy is a person's belief about their capabilities to perform towards a specific goal [9]. High self-efficacy improves one's ability to learn [22], perform tasks [9], exert agency and persist [43], and find enjoyment in challenges [17]. People with strong self-efficacy respond less negatively to failure and focus on strengths [18]. Critique, setback, and risks make creative work extremely challenging [52], and high self-efficacy provides an important robustness. With low self-efficacy, people are more likely to construe critique as an assessment of them, rather than as an assessment of the concept [39]. Recognizing this, the studio model of art and design education emphasizes critiquing *the work*, rather than the person [50].

Tohidi *et al.* revealed that potential users of interactive systems withhold critique when presented with a single prototype; the users were concerned about offending the designer [58]. More importantly, Tohidi et al. showed that the presence of multiple alternative concepts gave users license to be more critical with their comments. This paper explores the other side of the coin: how critiquing designer-generated alternatives affects the *designer's* self-efficacy. This paper hypothesizes that parallel prototyping changes the investment mindset: it encourages investment in a creative process rather than in a particular idea. Serial prototyping may lead people to fixate on a single concept, causing them to construe critique as a rebuke of their only option.

Hypothesis 3: *Parallel prototyping leads to a greater increase in design task-specific self-efficacy.*

In the ad study, self-efficacy is measured with a multi-question self-report assessment, administered before and after the design task.

METHOD

The study described in this paper manipulates the structure of the prototyping process. Web advertising was chosen because it fulfilled the following criteria:

- Quality can be measured objectively and subjectively;
- Participants need minimal artistic or engineering ability;
- Individuals can complete tasks within a single lab session;
- Solutions demonstrate creative diversity and a range of performance quality;
- The study procedure could generate consistent and useful feedback during iteration.

Study Design

The experiment employed a between-subjects design with one independent variable: the *structure* of the prototyping process. The study held constant the number of prototypes created, the amount of feedback provided, and the overall time allotted. In the *Parallel* condition, participants created 3 prototypes and then got feedback, then made 2 more and got more feedback, then a final version. In the *Serial* condition, participants create 5 prototypes in series, receiving feedback after each prototype, then a final version. Parallel participants were instructed to start subsequent prototypes at the same intervals as serial participants.

Participants

Thirty-three participants were recruited with fliers and assigned to one of two conditions. (Of 36 recruited, 3 dropped out before the end.) Participants' average age was 22; three-fourths were students. Using a stratified randomization approach, the study balanced gender (19 females) and prior design experience across conditions. Fourteen participants reported some prior experience in ad or graphic design; none were professional designers.

Materials

Graphic design tool

Participants designed a 160 × 600 pixel banner advertisement to be hosted on the social networking site MySpace.com. Ads were created using MySpace's Flash-based *AdBuilder* tool (see Figure 2). This simple graphic design tool was easy to learn, and no participants had used it before. Selecting a novel tool removes the confound of fluency with particular software. To insure a base level of competence, all participants had to successfully replicate a sample graphic in less than ten minutes.

Advertising client

Participants all created ads for the same client, *Ambidextrous* magazine [1], a student-led design publication. A design brief described the magazine's purpose and the kind of advertising desired by the client (Appendix D).

Figure 2 The ad design study used MySpace's *AdBuilder,* a browser-based graphic design tool

Prototype critique system

Prior to the experiment, a team of three advertising and graphic design professionals developed a list of about 50 statements that could serve as critique for banner ads (see Appendix A). The list included three categories of statements—overall theme, composition & layout, and surface elements. Each category contained 12 to 20 statements, intended to provide high-level direction, without using explicitly positive or negative language. These statements express basic graphic design principles. During the study, the experimenter chose three statements—one from each category—to attach to each ad prototype (see Figure 3).

The experimenter chose critiques relevant to each prototype and never repeated statements for the same participant. This process was identical for both conditions. Neither condition explicitly compared a participant's ads, such as, "The color in this ad is better than that one." In parallel, the experimenter reviewed each ad sequentially so that the process was equivalent in both conditions. The discussion section provides an in-depth treatment of the potential for bias in the study critique system.

Dependent Measures

Performance

After the experiment, all 33 final ad designs were uploaded

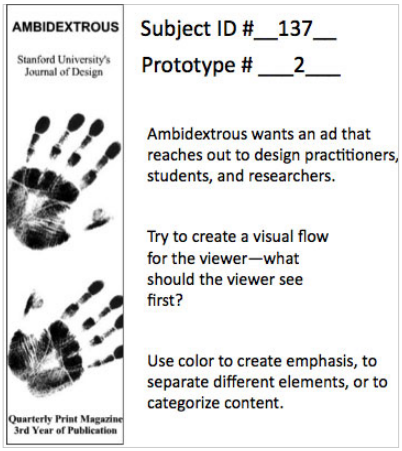

Figure 3 Example critique

to MySpace for a 15-day campaign targeted to users interested in design-related activities. This study's total advertising costs were under $200. Design performance was determined through two objective measures:

- MySpace click-through rates (CTR): daily number of clicks divided by the number of impressions (number of appearances on MySpace), and
- Google Analytics [2] on the target client Website: number of visitors, time spent, and number of pages visited daily from each ad.

Moreover, ads were independently judged by the magazine editors and by ad professionals. Editorial staff and ad professionals represent two important—and different—stakeholder perspectives. Four magazine editors and three advertising professionals rated the participants' ad designs from 0 to 10 along five dimensions: adherence to the client's theme, creativity/originality, visual appeal, tastefulness, and adherence to graphic design principles. Raters were blind to condition and rated ads individually, with no knowledge of other raters' scores.

Divergence

Creating a diverse set of ideas helps people understand the space of designs and their relative merits [12]. To obtain a measure of idea diversity, independent raters assessed pair-wise similarity of all combinations of each participant's ads (see Figure 4). Raters were recruited from Amazon Mechanical Turk, a crowdsourcing system for paying workers for short online tasks [3]. For each ad, raters assessed similarity on a scale from 1 to 7 (not similar to very similar).

Figure 4 Example pair-wise ad similarity rating, a measure of design divergence

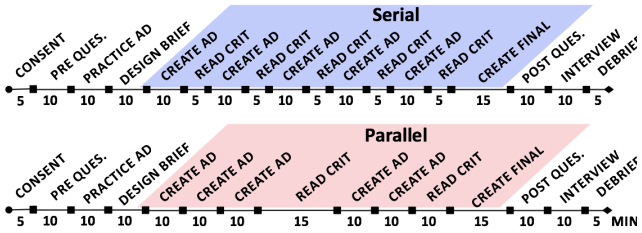

Figure 5 Procedure for Serial and Parallel conditions, with timing

Each rater assessed a randomly ordered set of at least 50 ads. Rating a large number of ads helped raters calibrate their assessments. This measure generated 14,850 judgments (30 worker assessments on each of the 15 pair-wise comparisons for 33 participants).

Self-efficacy

Questions on self-efficacy assessed participants' views of their graphic design ability (adopted from self-efficacy assessments in education [24,31]). The assessment asks participants to rate their ability to: create advertisements, understand design problems, detect problems in a design idea, and incorporate feedback into a design idea (see Appendix B). Each question solicited a 7-point Likert scale response. The same questions were administered before and after the design task, creating a difference measure (the time between the pre and post test was 1¾ hours). Comparing the change in self-efficacy measures how the process manipulation (Parallel/Serial) influenced an individuals' belief in their design abilities.

Procedure

The experiment had the following steps (see Figure 5): consent form, pre-task questions and demographics, practice task, design brief/instructions, prototyping periods (10 minutes per prototype), critique reports (5 minutes per prototype), final design period (15 minutes), post-task questions, an open-ended interview, and a final debriefing to reiterate the consent details. The practice task required participants to replicate a graphic (unrelated to the main task). The design brief detailed the ad campaign's client, *Ambidextrous* magazine and outlined three goals: increase traffic to the *Ambidextrous* Web site, impress the editors, and create ads with effective graphic design.

Participants were instructed they would receive critique from an ad expert on each prototype. As experimenters prepared critique reports in a separate room, participants were allowed to navigate the client Web site, search for images, or sketch on paper. After a set amount of time (two minutes per ad), participants received an envelope containing the printed ad prototype with feedback statements. As part of the final questionnaire, participants filled out the "Creativity Achievement Questionnaire" developed by Carson et al. to assess creative achievement across ten domains (visual arts, music, dance, architecture, writing, humor, inventions, scientific discovery, theater, and culinary arts) [13].

For 150 minutes of participation, subjects received $30 cash. Experiment proctors only entered the participant room

to introduce the tool and task, to deliver feedback envelopes, and to conduct the open-ended interview.

RESULTS

Participants generated a wide variety of ad concepts. The most successful ads (high click-through rates and ratings) tended to be simple, visually balanced, professional, creative, matched the theme of the magazine and contained some sort of intriguing hook, such as the face made of hands in the highest click-through performer (see Figure 6).

The study supported all three hypotheses. Participants in the Parallel condition produced higher quality designs (better click-through rates and higher subjective ratings) and more divergent prototypes. They also reported a greater increase in task-specific self-efficacy. Participants with prior experience in ad or graphic design outperformed complete novices, however the prototypes created by experienced participants were less diverse than novices.

Parallel Ads Outperformed Serial Ads

Online Click-Through Rates

Performance data on each ad was extracted from MySpace and Google Analytics on the *Ambidextrous* Web site (see Table 1). MySpace reports that over the 15-day campaign, the 33 participant ads received 501 total clicks on 1,180,320 total impressions (*i.e.,* number of ad appearances), giving an overall average click-through rate (CTR) of 0.0424% or 424 clicks per million impressions. The top two click-through rates were both Parallel ads, with 735 and 578 clicks per million impressions, respectively. The bottom two ads were both from the Serial condition; neither received any clicks.

Figure 6 Example ads: (Left) Parallel ad, 1st in click-through rate, 6th in expert rating; (Middle), Parallel ad, 9th in CTR, 1st in expert rating; (Right) Serial ad, 4th in CTR, 32nd in expert rating.

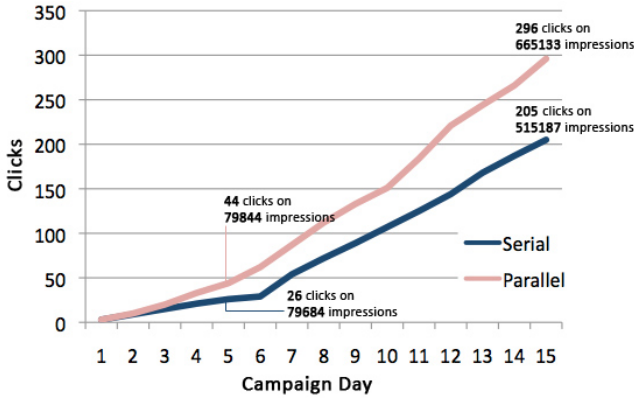

Figure 7 Parallel ads received more clicks—and more clicks per impression—than serial ads during a 15-day campaign.

MySpace users clicked Parallel ads more than Serial ads. Counting clicks can be misleading because some ads are shown more than others: when an ad performs well, the host often shows it more[1]. There are two approaches for measuring performance comparably. The first is to measure clicks per impression. The second is to hold impressions constant and compare clicks. In this study, ads received an approximately equal number of impressions for the first five days. A chi-squared analysis examines performance through day five. Parallel ads had 79,800 impressions with 44 clicks and Serial ads had 79,658 impressions with 26 clicks (see Figure 7); at this early stage, Parallel ads had a significantly higher click-through rate (χ^2= 4.60, $p<0.05$).

Over the entire campaign, an analysis of variances was performed with condition (Serial/Parallel) and Creativity test scores (high/low) as factors and final click-through rates for each ad as dependent variable. Parallel outperformed Serial, 445.0 and 397.9 clicks per million impressions respectively ($F(1,30)=4.227$, $p<0.05$) (see Table 1)[2]. Also, high Creativity scorers had a higher average click-through rate (352 clicks per million) than low scorers (305); this difference is not significant ($F(1,30)=3.812$, $p=0.06$).

---

[1] Like many advertising hosts, MySpace varies the number of impressions based on prior performance of the ad. MySpace does not publish their algorithm for determining the frequency of impressions, but a repeated measures general linear model with the Day 5 CTR as a factor and impressions on each subsequent day as dependent measure shows the CTR for days 1-5 to be a significant predictor of the number of impressions for the final 10 days of the campaign ($F(1,29)=23.2$ and $p<0.01$). MySpace receives payment on each click; intuitively, it is in their interest to show high-CTR ads more often.

[2] According to Google Analytics, the Ambidextrous Web site received 422 total visitors during the 15-day campaign, 79 less than the number of clicks reported by MySpace. One explanation for the disparity could be that users clicked the ad and then hit "back" before the browser loaded the client site. The 264 visitors from Parallel ads are significantly more than the 158 visitors from Serial when compared against impressions ($\chi 2$= 6.61, $p<0.05$).

Visitor Behavior on Client Site

One common measure of ad effectiveness is time on site [56]. The average time on site for Parallel ads (31.3 seconds) was greater than Serial ads (12.9 seconds) ($t(493)=1.781$, $p<0.05$). The result suggests that Parallel ads were more likely to reach people genuinely interested in the product offered by the clients. The number of pages visited per visitor was about the same: 1.49 for Parallel and 1.25 for Serial. Visitor's navigation behavior did not show a statistical difference: 71 of 264 visitors from Parallel ads and 35 of 158 visitors from Serial ads visited pages beyond the front page of *Ambidextrous*' website ($\chi^2=1.18$, $p>0.05$).

Independent Expert Ratings

Overall rating contained five 10-point rating scales: adherence to the client's theme, creativity/originality, visual appeal, tastefulness, and adherence to graphic design principles. The average expert rating across all ads was 23.0 out of 50 (35.6 high and 15.0 low). The three top-rated ads were all from the Parallel condition. An analysis of variances was performed with condition (Parallel/Serial), prior design experience (some/none), rater (seven independent raters), and rater type (client or professional) as factors and overall rating as the dependent variable. Parallel ads were rated higher ($\mu=24.4$, $SD=9.7$) than Serial ads ($\mu=21.7$, $SD=8.8$) ($F(1,203)=3.871$, $p<0.05$) (see Figure 8). Experienced participants created higher-rated ads ($\mu=25.7$, $SD=9.6$) than novices ($\mu=21.0$, $SD=8.6$) ($F(1,203)=20.98$, $p<0.05$) (see Figure 10). There was no interaction effect between condition and prior experience.

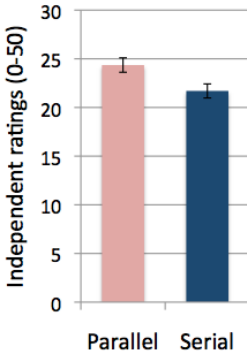

Figure 8 Parallel ads received higher average independent ratings (0-50 scale) than serial ads

Some raters had higher average ratings than others ($F(5,203)=18.88$, $p<0.05$). There was no interaction between rater and condition; raters generally agreed that parallel ads outperformed serial ads. Analyses of variances were conducted separately for all five dimensions with condition and experience as factors. All dimensions skewed towards Parallel ads, but only two—tastefulness ($F(1,227)=7.527$, $p<0.05$) and adherence to graphic design principles ($F(1,227)=4.188$, $p<0.05$)—were independently significant in favor of Parallel ads. The ratings provided by the clients were higher on average ($\mu=24.3$, $SD=9.5$) than those provided by external ad professionals ($\mu=22.0$, $SD=9.1$) ($F(1,203)=4.376$, $p<0.05$). There was no interaction effect between rater type and condition.

Ads that performed well online generally also received high ratings by the clients and ad professionals. The ad with the best overall click-through rate received the 6th highest rating by the clients and ad professionals (see Figure 6, left). Likewise, the highest rated ad achieved the 4th highest click-through performance (see Figure 6, middle). There were anomalies, such as the top two ads in the Serial condition. These two ads were ranked 25th and 32nd (out of 33) by the expert raters, but received the 3rd and 4th best overall click-through rates. The latter of those designs does not even mention the client (see Figure 6, right). Statistically speaking, online click performance was not a predictor of overall expert rating ($R^2 = 0.057$, $F(1,31) = 1.858$, $p > 0.05$, $b = 0.192$).

	Parallel	Serial
Performance data from advertising host (MySpace.com)		
Total impressions	665,133 (43968)	515,187 (36373)
Total clicks	296 (22.8)	205 (19.1)
Clicks per million impressions	445.0 (18.3)	397.9 (19.6)
Performance data on client site (Google Analytics reports)		
Total visitors	264 (19.9)	158 (15.3)
Average time (sec) per visitor	31.3 (143)	12.9 (79.9)
Pages visited on site	394 (31.6)	198 (21.1)
Pages visited per visitor	1.49 (0.48)	1.25 (0.41)

Table 1 Summary of campaign data from MySpace and Google Analytics (standard deviation in parentheses).

Parallel Ads Were Rated More Diverse Than Serial Ads

Workers on Amazon Mechanical Turk rated Parallel ads as more divergent than Serial ads. Raters performed pair-wise similarity comparisons on a scale of 0 to 7 within each participant's set of six prototype ads. An analysis of variances was performed with condition (Serial/Parallel) and prior design experience as factors and pair-wise similarity rating as the dependent variable. Serial ads were deemed significantly *more similar* ($\mu=3.25$, $SD=1.96$) than Parallel ads ($\mu=2.78$, $SD=1.66$) ($F(1,14816)=239.3$, $p<0.05$). Parallel ads were rated more divergent[3].

Similarity ratings were not predictive (or inversely predictive) of online click performance ($R^2 = 0.032$, $F(1,31) = 0.030$, $p > 0.05$, $b = 0.009$) or overall independent ratings ($R^2 = 0.030$, $F(1,31) = 1.999$, $p > 0.05$, $b = 0.246$).

Parallel Participants' Ad Design Self-Efficacy Increased

A self-efficacy assessment measured participants' belief in their ability to perform the design task. The difference between the pre- and post-task scores provides an indication of how participants' beliefs change. Across all participants, self-efficacy rose from 10.85 to 12.12 (out of 20); a paired-

[3] Similarity ratings changed depending on whether they were generated early or late in the process. Pair-wise comparison of pairs 1-2, 2-3, and 1-3 were labeled "Early" designs; pairs 4-5, 5-6, and 4-6 were labeled "Late" designs. An analysis of variances was performed with condition (Serial/Parallel) and design-stage pairs (Early/Late/Other) as factors and similarity rating as the dependent variable. Across conditions, ads created *later* were deemed more similar ($\mu=3.41$, $SD=2.03$) than *early* ads ($\mu=2.97$, $SD=1.77$) ($F(1,14814)=107.835$, $p<0.05$). The interaction between condition and design stage was marginally significant ($F(1,14814)=2.460$, $p=0.085$). Serial ads were rated more similar than Parallel ads, both for early and late pairs, but the similarity is greater for later ads.

samples T-test shows a significant difference ($t(32)=2.355$, $p<0.05$). Examining inter-question effects, each question independently resulted in a significant rise from pre to post-task, except for question four ("rate your ability to incorporate feedback into a design idea") ($t(32)=0.154$, $p>0.05$). This rise is consistent with prior findings that show individual self-efficacy beliefs increase with practice [9,31].

An analysis of variances was performed with condition (Serial/Parallel) and prior task experience (Experienced/Novice). Participants in the Parallel condition reported a significant increase in self-efficacy scores (see Figure 9), a net gain of 2.5 points ($F(1,29)=4.210$, $p<0.05$), while the Serial condition essentially remained even (net gain $\mu= 0.4$).

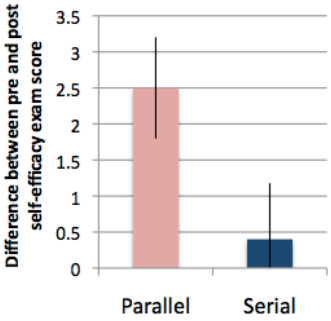

Figure 9 Participants in the Parallel condition reported a greater increase in self-efficacy from pre to post design task

Experienced Participants Outperformed Novices

Online Click-Through Rates

The fourteen participants with prior experience in ad or graphic design significantly outperformed novices. Ads by participants with prior experience received 350 clicks on 752,424 impressions, compared to 151 clicks on 427,896 impressions by novices ($\chi^2= 8.10$, $p<0.05$). There was no interaction effect between condition and prior participant experience.

Visitor Behavior on Client Site

Visitors spent more time on the client's site after clicking ads created by experienced participants (38.0 sec/visitor) compared to those created by novices (7.6 sec/visitor) ($F(1,491)=8.456$, $p<0.05$). An interaction between condition and prior experience showed that having prior experience in the Parallel condition led to more time on site than prior experience in the Serial condition, 57.0 to 18.9 seconds/visitor ($F(1,491)=4.045$, $p<0.05$). Visitors from experienced ads were also more active navigators; 88 of 296 visitors from experienced ads and 12 of 126 visitors from novice ads visited pages beyond the front page of *Ambidextrous*' website ($\chi^2=19.96$, $p<0.05$).

Divergence Ratings

Participants with prior experience created significantly *more similar* ads ($\mu=3.15$, $SD=1.86$) than novices ($\mu=2.88$, $SD=1.80$) ($F(1,14816)=76.70$, $p<0.05$). Ads created by novices were rated more divergent. There was also an interaction effect indicating that experienced participants in the serial condition created the most similar ads ($F(1,14816)=36.45$, $p<0.05$).

Self-Efficacy Assessment

Participants with prior ad design experience reported a similar gain in self-efficacy ($\mu=1.93$) as novices ($\mu=0.79$) ($F(1,29)=1.088$, $p>0.05$). There was an interaction effect between condition and prior experience: novices reported a 2.9 increase in self-efficacy in Parallel, but a 0.73 decrease in Serial ($F(1,29)=6.331$, $p<0.05$) (see Figure 10). In short, parallel prototyping positive affected an individual's belief in their ad design ability, especially for novices.

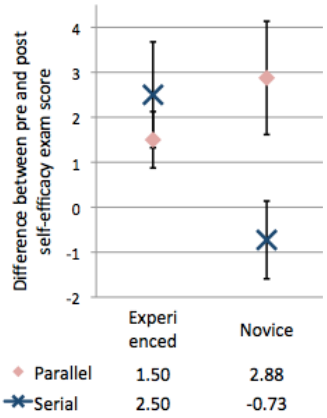

Figure 10 Novice participants in the Parallel condition reported an increase in self-efficacy from pre to post design task; self-efficacy for novices in serial decreased

ANALYSIS

A parallel approach led to ad creations with better performance by every measure: higher independent ratings, more impressions served up by MySpace, better click-through rates, more visitors to the client Web site, and more site interaction per visitor. Participants created the same number of prototypes and received equivalent feedback in the same time period. The only difference between conditions was a matter of *when* participants received critique on their ideas—after each concept or after multiple creations.

Why did the process manipulation affect performance? This section offers three explanations for the differential results: comparison helped the parallel participants learn ad design principles, parallel participants better explored the design space, and serial participants perceived the critique as negative and thus gained no confidence at ad design.

Did parallel feedback impact how participants learned? Comparison processes can facilitate inductive reasoning on rival observations [14,57]. Since Parallel participants received feedback on multiple ideas simultaneously, they were more likely to read and analyze critique statements side-by-side. Direct comparison perhaps helped Parallel participants better understand key design principles and lead to more principled choices for subsequent prototypes. In Serial prototyping, participants' ideas tended to follow directly from the feedback. This serial approach may implicitly encourage refinement at the expense of exploration. Performance likely improves in Parallel because people exercise their comparative abilities to learn contextual constraints, malleable variables, and their interrelations.

Learning a Parallel approach may change future behavior. When asked to describe their process for future design projects, 11 of 16 Parallel participants said they would create more than one prototype and obtain copious feedback; only 5 of 17 Serial participants made similar claims ($\chi^2= 2.63$, $p>0.05$). As one Parallel participant said, "not spending too much time on any single prototype is useful because then you don't go into details too much."

Did a parallel process impact how participants explored concepts? The study showed Parallel participants created significantly more divergent prototypes; Serial participants tended to create more similar designs. The interviews re-

vealed the role of critique, as one Serial participant explained, "I think the feedback helped. I kept repeating the same mistakes, but maybe less and less each time… the feedback reiterated that." Another Serial participant said:

> I would try to find a good idea, and then use that idea and keep improving it and getting feedback. So I pretty much stuck with the same idea.

This notion of "sticking" with an idea or using the feedback to decide where to go next did not surface in the Parallel condition. As one Parallel participants reported: "I didn't really try to copy off of the ads that I did before...I just made new ideas." Both the divergence measure and the qualitative interviews suggest the parallel structure supports more generative thinking and reduces fixation.

Parallel prototyping may encourage both a broad enumeration stage and a subsequent reflection stage. By contrast, Serial's immediate feedback implicitly encourages refinement. On this view, the fact that Parallel *delays* feedback is actually an advantage. From a behaviorist perspective, this can seem counterintuitive because immediate feedback highlights the connection between cause and effect. However, delay helps learners *reflect*: readily available, immediate feedback can be a crutch that discourages building a deep understanding [5,49].

There are countless ways to combine text, images, and backgrounds in a 160×600 pixel ad design; some combinations perform better than others. To use an analogy, exploring design possibilities is like simulated annealing [30]. Creative work often benefits from broadly exploring a design space with high entropy before optimizing in one direction. Perhaps serial participants hill-climbed to local, rather than global optima.

Experienced participants created a less diverse set of designs than novices; they also outperformed novices. In general, experts may know a priori which areas are promising and which to avoid. By contrast, novices have to learn what is effective through trial and error.

Did Parallel participants gain more confidence in their ad-design ability? Parallel participants reported self-efficacy gains, while the Serial participants reported no change. This effect was more pronounced for novices. Serial participants also perceived the expert feedback more negatively. In open-ended interviews, 13 of 16 Parallel participants said the feedback was helpful or intuitive compared to 6 of 17 in Serial (χ^2= 7.13, $p<0.05$). More notably, 8 of 17 of the Serial participants reported the feedback as negative, compared to no such reports in the Parallel condition (χ^2=9.94, $p<0.05$). One participant in the Serial condition said:

> I received really negative comments saying [the clients] are looking for a creative and clever ad, which in other words is saying that this is stupid or ridiculous.

Moreover, participants were asked to leave their email if they wanted to later volunteer for *Ambidextrous* magazine. Twelve out of sixteen Parallel participants provided their email, while only five of seventeen did the same in Serial (χ^2= 6.86, $p<0.05$), which suggests the Parallel process may have helped motivate future action.

Perhaps having multiple alternative designs encourages investment in a creative process rather than a particular idea. Consequently, the parallel process encourages viewing critique as an opportunity for improvement. In contrast, the fixation engendered by serial prototyping may cause people to take critique as a catastrophic rebuke of their only option. With only one option there is no separation between designer and design. Parallel offers people distance between ego and object; Serial conflates them.

FOLLOW-UP STUDIES

Two short follow-up experiments examined questions raised by the main study.

Did the experimenters (possibly subconsciously) provide better critique to the parallel participants? To assess bias, two ad professionals unfamiliar with the experimental manipulation provided blind-to-condition independent assessments of the critique statements. The expert judges performed a selection task resembling the task performed by the experimenter. After reading about the client's advertising needs, the judge viewed an ad prototype and two triads of critique statements; one triad contained the three statements chosen during the actual experiment and the other triad was a random selection from the critique statement corpus. Judges were instructed to select a triad that "provides the most appropriate critique for the advertisement."

An intra-class correlation (ICC) with a two-way mixed model [53] calculated the reliability between the experimenter's choice of statements and each expert judge's choice. The ICC(3,1) single measure correlation among raters on Parallel ads is 0.434, and 0.496 for Serial ads. There is no significant difference between these numbers and both represent a moderate level of inter-rater reliability [40]. In short, experts generally agreed with the feedback provided, and the level of agreement was comparable across conditions.

Did the critique statements help participants produce better ads? A follow-up study examined the value of the scripted ad critique statements in Appendix A. Thirty participants followed a serial process to create three prototypes and one final advertisement. The final ads were launched in an online ad campaign and rated by experts. Participants were randomly assigned to one of three conditions: *targeted*, *random*, and *none*. In the *targeted* condition, an experimenter selected three critique statements intended to help the participant improve their design. In the *random* condition, a random algorithm selected three critique statements. Participants in the *none* condition received no critique; rather, they viewed the client Web site during an equivalent critique interval.

In a 3-week campaign, ads that received targeted critique had 49,414 impressions with 203 clicks, ads with no feedback had 49,663 impressions with 179 clicks, and ads that received randomly selected critique statements received

49,543 impressions with 157 clicks ($\chi 2$= 6.01, $p<0.05$). Moreover, twenty independent experts rated ads (on a 0 to 30 scale) with targeted critique higher μ=15.9 (SD=5.4) than ads with random critique μ=15.1 (SD=5.2) and ads with no critique μ=14.4 (SD=6.2) ($F(2,597)=3.287$, $p<0.05$). The study found that targeted critique helped participants learn basic principles of graphics design and produce better ads.

DISCUSSION ABOUT EXPERIMENTAL APPROACH

The experimental paradigm introduced in this paper provides several important benefits for studying design. First, design results can be objectively measured through real-world analytics, and subjectively assessed through crowd-sourced and stakeholder ratings. Second, solutions demonstrate creative diversity and exhibit a broad range of performance. Third, it offers a mechanism for presenting feedback interactively and studying its effects. The advertising domain achieves theses goals particularly well. Hopefully this paradigm will prove useful in additional domains.

Web analytics can be tremendously valuable for experimental work; it also presents several challenges. Web hosts often show ads differentially based on performance. Poor performing ads must have a large enough number of impressions to yield a robust measure of click-through rate. Additionally, click-through rate can vary over time. Fair comparison requires holding the number of impression constant, analyzing data from a time interval with roughly balanced impression rate, or using more sophisticated statistical analysis to factor out time effects.

CONCLUSIONS AND FUTURE WORK

This paper found that when people create multiple alternatives in parallel they produce higher-quality, more-diverse work, *and* experience a greater increase in self-efficacy. Many excellent designers practice this approach already; their successes inspired this research. Hopefully, these results will encourage many more practitioners and teachers to adopt a parallel approach. Integrating the parallel approach into design practicum can inculcate healthy prototyping habits and help foster a positive outlook toward critique. In the future, software tools and infrastructure providers could provide a powerful benefit by enabling people to rapidly design alternatives and experimentally compare them. More broadly, this research seeks to develop a theoretical understanding of creative work to help practitioners and students solve design problems more effectively. An important direction for future work is to study the impact of parallel design in other contexts, especially at longer time scales and for design teams.

ACKNOWLEDGEMENTS

We thank Ugochi Acholonu, Lera Boroditsky, Jodi Forlizzi, Jeffrey Heer, Erik Stolterman, Ewart Thomas, and Terry Winograd for valuable discussions about this research. We thank Rachel Lopatin for analyzing the targeted critique data. The Hasso Plattner Design Thinking Research Program financially supported this work; Intel Corporation donated computers.

REFERENCES

1. Ambidextrous Magazine. http://ambidextrousmag.org/.
2. Google Analytics. http://www.google.com/analytics/.
3. Amazon Mechanical Turk. https://www.mturk.com/mturk/.
4. Alexander, C., Ishikawa, S., and Silverstein, M. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, 1977.
5. Anderson, J. and Schooler, L. The Disruptive Potential of Immediate Feedback. *CMU Department of Psychology*, (1990), 702-708.
6. Arkes, H.R. and Blumer, C. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes 35*, 1 (1985), 124-140.
7. Ball, L.J., Jonathan St.B.T., Evans Ian Dennis, and Ormerod, C. Problem-solving Strategies and Expertise in Engineering Design. *Thinking & Reasoning 3*, 4 (1997), 247.
8. Ball, L.J. and Ormerod, T.C. Structured and opportunistic processing in design: a critical discussion. *International Journal Human-Computer Studies 43*, 1 (1995), 131-151.
9. Bandura, A. *Self-Efficacy: The Exercise of Control*. Worth Publishers, 1997.
10. de Bono, E. *Six Thinking Hats*. Back Bay Books, 1999.
11. Boroditsky, L. Comparison and the development of knowledge. *Cognition 102*, 1 (2007), 118-128.
12. Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, 2007.
13. Carson, S.H., Peterson, J.B., and Higgens, D.M. Reliability, Validity, and Factor Structure of the Creative Achievement Questionnaire. *Creativity Research Journal 17*, 1 (2005), 37-50.
14. Colhoun, J., Gentner, D., and Loewenstein, J. Learning abstract principles through principle-case comparison. *Proceedings of Cognitive Science Society*, (2008), 1659-1664.
15. Cross, N. Expertise in design: an overview. *Design Studies 25*, 5 (2004), 427-441.
16. Cross, N. *Designerly Ways of Knowing*. Springer, 2006.
17. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*. Harper Perennial, 1991.
18. Dodgson, P. and Wood, J. Self-esteem and the cognitive accessibility of strengths and weaknesses after failure. *Journal of Personality and Social Psychology 75*, 1 (1998), 178-197.
19. Dorst, K. and Cross, N. Creativity in the design process: co-evolution of problem-solution. *Design Studies*

22, 5 (2001), 425-437.

20. Dow, S.P., Heddleston, K., and Klemmer, S.R. The Efficacy of Prototyping Under Time Constraints. *Proceeding of the seventh ACM conference on Creativity and cognition*, ACM (2009), 165-174.
21. Duncker, K. *On Problem-solving*. Greenwood Press Reprint, 1945.
22. Dweck, C. *Mindset: The New Psychology of Success*. Ballantine Books, 2007.
23. Dym, C.L., Agogino, A.M., Frey, D.D., and Leifer, L.J. Engineering Design Thinking, Teaching, and Learning. *Journal of Engineering Education 94*, (2005), 103-120.
24. Fredrickson, B.L. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist 56*, 3 (2001), 218-226.
25. Gentner, D., Loewenstein, J., and Thompson, L. Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology 95*, (2003), 408, 393.
26. Gentner, D. and Markman, A.B. Structure mapping in analogy and similarity. *AMERICAN PSYCHOLOGIST 52*, (1997), 45--56.
27. Gick, M.L. and Holyoak, K.J. Schema induction and analogical transfer. *Cognitive Psychology 15*, 1 (1983), 1-38.
28. Goel, V. and Pirolli, P. The structure of design problem spaces. *Cognitive Science 16*, 3 (1992), 395-429.
29. Gopnik, A., Meltzoff, A.N., and Kuhl, P.K. *The Scientist in the Crib: What Early Learning Tells Us About the Mind*. Harper Paperbacks, 2001.
30. Granville, &. Simulated annealing: A proof of convergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 6 (1994), 652–656.
31. Hall, T.S. Improving Self-Efficacy in Problem Solving: Learning from Errors and Feedback. 2008.
32. Hartmann, B., Klemmer, S.R., Bernstein, M., et al. Reflective physical prototyping through integrated design, test, and analysis. *Proceedings of the 19th annual ACM symposium on User interface software and technology*, ACM (2006), 299-308.
33. Herring, S.R., Chang, C., Krantzler, J., and Bailey, B.P. Getting inspired!: understanding how and why examples are used in creative design practice. *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 87-96.
34. Janis, I.L. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Wadsworth Publishing, 1982.
35. Jansson, D. and Smith, S. Design Fixation. *Design Studies 12*, 1 (1991), 3-11.
36. Kelley, T. *The Art of Innovation*. Broadway Business, 2002.
37. Kershaw, T.C. and Ohlsson, S. Multiple causes of difficulty in insight: the case of the nine-dot problem. *Journal of Experimental Psychology. Learning, Memory, and Cognition 30*, 1 (2004), 3-13.
38. Kolodner, J. *Case-Based Reasoning*. Morgan Kaufmann, 1993.
39. Kosara, R. Visualization Criticism - The Missing Link Between Information Visualization and Art. *Proceedings of the 11th International Conference Information Visualization*, IEEE Computer Society (2007), 631-636.
40. Landis, J.R. and Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics 33*, 1 (1977), 159-174.
41. Laseau, P. *Graphic Thinking for Architects and Designers, 2nd Edition*. John Wiley & Sons Inc, 1988.
42. Lee, B., Srivastava, S., Kumar, R., Brafman, R., and Klemmer, S.R. Designing with Interactive Example Galleries. .
43. Mele, A.R. *Motivation and Agency*. Oxford University Press US, 2005.
44. Newell, A. *Human Problem Solving*. Prentice Hall, 1972.
45. Nielsen, J. and Faber, J.M. Improving System Usability Through Parallel Design. *Computer 29*, 2 (1996), 29-35.
46. Osborn, A.F. *Applied Imagination: Principles and Procedures of Creative Problem Solving*. Charles Scribner's Sons, 1963.
47. Piaget, J. *The Psychology of Intelligence*. Routledge, 2001.
48. Rittel, H. and Webber, M. Dilemmas in a general theory of planning. *Policy Sciences 4*, 2 (1973), 169, 155.
49. Schmidt, R., Young, D., Swinnen, S., and Shapiro, D. Summary Knowledge of Results for Skill Acquisition: Support for the Guidance Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition 15*, 2 (1989), 352-359.
50. Schon, D.A. *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. Jossey-Bass, 1990.
51. Schon, D.A. *The Reflective Practitioner: How Professionals Think in Action*. Ashgate Publishing, 1995.
52. Schrage, M. *Serious Play: How the World's Best Companies Simulate to Innovate*. Harvard Business School Press, 1999.
53. Shrout, P.E. and Fleiss, J.L. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin 86*, 2 (1979), 420-428.

54. Simon, H.A. *The Sciences of the Artificial - 3rd Edition*. The MIT Press, 1996.

55. Smith, S., Kohn, N., and Shah, J. What You See Is What You Get: Effects of Provocative Stimuli on Creative Invention. *NSF International Workshop on Studying Design Creativity*, (2008).

56. Sterne, J. *Web Metrics: Proven Methods for Measuring Web Site Success*. Wiley, 2002.

57. Thompson, L., Gentner, D., and Loewenstein, J. Avoiding Missed Opportunities in Managerial Life: Analogical Training More Powerful Than Individual Case Training. *Organizational Behavior and Human Decision Processes 82*, 1 (2000), 60-75.

58. Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. Getting the right design and the design right. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 1243-1252.

59. Torrance, E.P. *Torrance tests of creative thinking*. Personnel Press, Ginn and Co., Xerox Education Co, 1974.

APPENDIX A: Expert Critique Statements

1. Overall/ Thematic

Ambidextrous seeks an ad with a single clear message that matches the theme of their journal.

Ambidextrous wants an ad that clarifies the product: a journal about design and design process.

Ambidextrous desires an ad that is simple, readable, consistent, and deliberate.

Ambidextrous does not want the ad to sound exclusive; they are open to anyone with interest.

Ambidextrous is looking for a creative and clever ad.

Ambidextrous is looking for a professional and tasteful ad.

Ambidextrous wants an exciting and visually appealing ad.

Ambidextrous wants an ad that matches the journal's style.

Ambidextrous wants an ad that reaches out to design practitioners, students, and researchers.

Use graphics/images that support the overall message. What message are you trying to convey?

Use colors/fonts that support the overall message. What message are you trying to convey?

Remember that the ad is a link; the URL does not *necessarily* have to be on the ad design.

2. Composition & Layout

Visual Flow and Balance

Try to create a balanced layout where the graphics don't tilt to one side or the other.

Try to create a visual flow for the viewer—what should the viewer see first?

Think about the proximity of different elements. How close together or far apart elements are placed suggests a relationship (or lack thereof) between otherwise disparate parts.

To help balance the ad, leave slightly more space at the bottom relative to the top of the ad.

Contrast the position of elements to draw the viewer's attention to the most important parts.

To create consistency for the viewer, create a consistent and balanced look using repetition.

Spacing and Alignment

Align text and graphics to create more interesting, dynamic, and appropriate layouts.

Use alignment to create a clean and organized look.

It's ok to break alignment only to draw the viewer's attention to important elements in the ad.

Use white around text and images to help frame the content.

Use space—the absence of text and graphics—to provide visual breathing room for the eye.

Try to balance the spacing around the border of the ad design.

These visual elements in the ad don't line up.

Consider playing around with different ways to justify the text (e.g., center, left, or right-justified).

Emphasis & Hierarchy

Be conscious of competing elements in the ad. Think about what should have emphasis.

Draw the viewer's attention to elements by contrasting size (scale).

Think about the visual hierarchy of the different elements (texts, images, colors, etc) of the ad. What is the most important?

Help the viewer recognize, identify and comprehend the most important information in the ad.

Use elements with visual intensity or color for emphasis.

3. Fonts, Colors, Images

Font Type

Try not to distort the font so that it becomes hard to read.

Use large, bold font/graphics to create focus or emphasis on the ad design.

If using text over an image, make the text bigger and darker than normal; make sure it is readable.

For text to stand out it has to be substantially different than other text.

Try not to mix serif and sans serif fonts.

Avoid using two different fonts that are too similar.

Try not to over emphasize text elements. (ex. a font does not need to be large, bold, and italic).

Images

Use large, bold graphics to create the focus of the ad design.

Consider using images for more visual impact.

Consider using fewer images.

Try not to over-rotate images, as it often distorts the content.

Color

Use color to create emphasis, to separate different elements, or to categorize content.

Avoid really light, bright colors.

Avoid colors together that look too similar (ex. brown & grey).

Try to use different colors that go well together.

Avoid complicated backgrounds.

Try to create a good visual separation between the text and the background

APPENDIX B: Self-Efficacy Questions (Pre & Post Task)

One a scale from "Not confident at all" (1) to "Very confident" (7), how confident are you:

1) With your ability to design advertisements?
2) At understanding design problems?
3) With detecting problems in your design?
4) With incorporating expert feedback in your design?

APPENDIX C: Post Interview Guide

These questions provided guidance for the final interview; the exact order and phrasing varied.

- Please describe the final design you came up with.
- What do you think are the strengths and weaknesses of your ad design?
- Describe your design process and why you created each design.
- How did the feedback affect you? Was it helpful? What did you learn about graphic ad design?

- If you created another ad, how would you approach it? Describe your design process. Would you seek feedback? How many prototypes would you create?

APPENDIX D: Advertising Design Brief

Assignment

You have been hired to design a graphic advertisement for *Ambidextrous*, Stanford University's Journal of Design. You will learn a new graphic design tool, prototype a number of example ads, receive feedback from an ad design expert, and then create a final ad for MySpace.com.

Goals

Keep in mind the following goals as you create your ads:

a) Increase traffic to the *Ambidextrous* website: http://ambidextrousmag.org/
b) Reach out to the target audience: designers, researchers, practitioners, and students who are interested in stories about the process of design.
c) Impress the editors of *Ambidextrous*. The client wants an ad that fits their overall aesthetic and theme (see below).
d) Create ads with effective graphic design.

What is Ambidextrous?

Ambidextrous is Stanford University's Journal of Design. In its 3rd year of publication, *Ambidextrous* is a quarterly subscription-based print magazine that features people and processes involved in design and design thinking. *Ambidextrous* is a forum for the cross-disciplinary, cross-market community of people with an academic, professional and personal interest in design. Each issue focuses on one general topic. Previous topics have included Space, Secrecy, Food, The Future, Danger, Developing, Borders & Interfaces, etc. Articles are written by the community at large, and edited, illustrated, and photographed entirely by volunteers.

Theme and Aesthetic for the Ambidextrous Ad

The *Ambidextrous* editors would like an ad that embodies the theme and general aesthetic of the journal. The journal tells stories about people who do design and the process of getting there, not just final products. Readers of the journal are not an exclusive club—it's intended to be accessible to folks without formal design training. In general they are looking for an ad that is tasteful, creative, professional, visually appealing, and conveys a clear message about the product.

Rules/Requirements

- You may download and use graphics, images, text etc. as you see fit.
- You may not use another company's logo, copyrighted images, profanity, obscenity or nudity. Unacceptable ads will be rejected by the research team.
- Do not include the magazine's URL on the ad. Clicking the ad will direct the user to the site.