

Prototyping Dynamics: Sharing Multiple Designs Improves Exploration, Group Rapport, and Results

Steven P. Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel L. Schwartz, Scott R. Klemmer

Stanford HCI Group

[spdown, srk]@stanford.edu

ABSTRACT

Prototypes ground group communication and facilitate decision making. However, overly investing in a single prototype can lead to fixation and impede the collaborative process. Does sharing multiple prototypes improve collaborative design? In a study, participants prototyped advertising designs alone and then discussed the campaign with a partner. In the *Share Multiple* condition, participants designed and shared three ads. In the *Share Best* condition, participants designed three ads and selected one to share. In the *Share One* condition, participants designed and shared one ad. Each condition was allotted the same time for design. In all conditions, partners critiqued each other's ads, then individually created a final ad. *Share Multiple* ads significantly outperformed other ads. Participants who designed and shared multiple prototypes integrated more of their partner's ideas into their own subsequent designs, explored a more divergent set of ideas, and provided more productive critiques of their partner's designs. Understanding how and why prototyping practices differentially affect results can have a broad impact on design research and education.

Author Keywords

Prototyping, group dynamics, iteration, feedback, comparison, divergence, exploration, critique, conceptual blending, design communication, group rapport

ACM Classification Keywords

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

General Terms

Experimentation, Design

INTRODUCTION

Many designers live by the principle, “never go to a client meeting without a prototype” [52]. Prototypes help people summarize their ideas, demonstrate progress and expertise, surface implicit design vocabulary, and ground group communication and decision making [13,44,45]. Sharing prototypes may also carry risks. First, the presence of a concrete prototype may (for better and for worse) focus the discussion on refining *that* idea rather than thinking more broadly [14,31]. Second, those presenting designs often believe their status to be on the line [13]. This risk encour-

ages overinvesting time, labor, psychological energy, and social momentum into a single concept [12,18]. In this single-prototype strategy, individuals may seek validation for their ideas and disregard or fear the critique and feedback necessary for exploration and revision [39]. Without exploration, people often interpret the frame of the design problem too narrowly [32]. Compounding this, collaborative work is often susceptible to *groupthink*, where members reinforce each other's belief in the current direction at the expense of other options [30].

An alternate strategy is for designers to share multiple rough prototypes during critiques instead of just one. This “parallel” strategy can help individuals more effectively understand underlying design principles, enumerate more diverse solutions, and react less negatively to feedback [18]. Distributing one's psychological investment across multiple prototypes can reduce fixation and sunk-cost reasoning [6,31]. Individuals may be more candid and critical of their own and others' ideas [53], resulting in more fluid and effective collaboration.

However, creating multiple alternatives leaves less time to polish each one and may be perceived as wasting effort [45]. Focusing on fewer endeavors can help people focus, contemplate, relax, and be more productive [29,36]. Increasing options can cause analysis paralysis—a “paradox” of choice [46]—and may jeopardize a group's ability to achieve consensus [9].

What are the impacts of preparing and sharing multiple designs? This paper investigates the hypothesis that sharing multiple prototypes increases design performance, improves group interaction, and leads to more effective idea sharing. In a between-subjects experiment, 84 participants working in pairs designed Web banner advertisements for a non-profit organization. The study comprised three steps. First, participants prototyped designs individually. Second, they worked with a partner to critique each other's ideas. Third, each individual created a final ad. Participants answered survey questions at several points and open-ended questions at the end. Pairs were randomly assigned to one of three conditions: creating and sharing multiple ads; creating multiple ads and sharing the best; and creating and sharing one ad. Comparing these three conditions separates the effects of *producing* multiple designs and *sharing* multiple designs.

The study found that ads in the *Share Multiple* condition generated more clicks per impression than the other conditions (see Figure 1). Independent (and blind-to-condition) judges rated ads from the *Share Multiple* condition signifi-

unpublished tech report, 09.24.2010

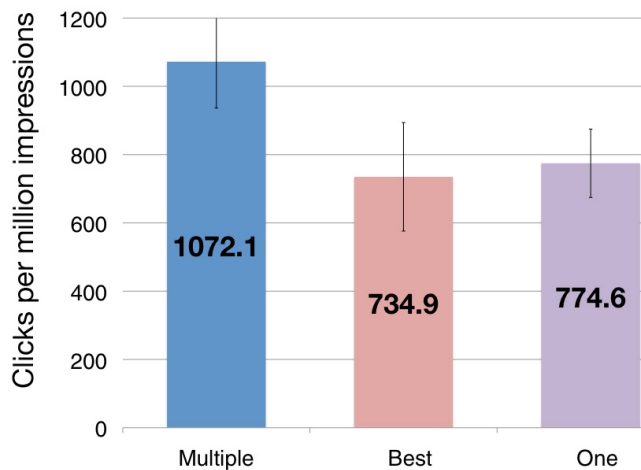


Figure 1 Online ad performance (clicks per million impressions): *Share Multiple* ads outperformed the other conditions.

cantly higher than the *Share One* and *Share Best* conditions. Judges also rated *Share Multiple* ads as significantly more divergent. Participants in the *Share Multiple* condition shared significantly more ideas and moved more towards consensus than pairs who shared only one design. Group members in the *Share Multiple* condition reported a greater increase in rapport over the course of the experiment, while rapport in the *Share One* and *Share Best* conditions dropped. Moreover, *Share Multiple* participants exchanged turns speaking significantly more often. Participants who scored high on a pretest of design principles created higher-rated ads.

In short, sharing multiple designs improves outcome, exploration, sharing, and group rapport. This result has significant implications for how designers and educators structure creative group work.

Exposure to examples enhances individual exploration

Exposing people to examples increases the likelihood they will integrate similar features into their own designs [42], even when they are asked to create vastly different ideas [22]. Furthermore, people's borrowing increases with the number of examples people see [37]. Smith *et al.* hypothesized that people often take the path of least resistance, and that this conformity constrains creativity [42]. However, using Smith *et al.*'s task, Marsh *et al.* found that participants who saw many examples created equally novel work [37]. In other words, participants borrowed from examples when they lacked a better idea, but viewing examples did not "push out" or inhibit people's novel ideas. Furthermore, when viewed from a *quality* perspective rather than a *novelty* perspective, people perform better when examples are readily available [11,34]. In all of this prior work, the examples were presented anonymously. Collaborative work is importantly different in this regard because the examples are produced by a known and co-present peer.

This paper hypothesizes that producing multiple prototypes and being exposed to multiple examples produced by other group members leads individuals to create a more divergent set of prototypes.

Hypothesis 1: Creating and viewing multiple prototypes leads to more individual design exploration.

This study measured individual design exploration by having independent raters judge the diversity/similarity of each participant's prototypes.

Sharing multiple prototypes improves collaboration

Designers often work collaboratively to generate, critique, and revise ideas, and to build consensus [25,44,55]. Under controlled conditions, individuals working separately often collectively produce a greater volume of ideas than group brainstorming [17,51]. Group members may block each other from sharing ideas [48], get frustrated with bad apples in the group [21], and deferentially follow others' ideas [30]. However, measuring only the sheer volume of ideas is misleading: group brainstorming supports organizational memory of design solutions, recognizes skill variety among team members, and builds shared ownership of ideas—crucial for selecting and refining concepts [50]. To some extent, the debate over whether to design individually or collectively presents a false choice; creative work typically involves both [10,49].

Sharing ideas with a group can be an anxiety-laden experience, and this anxiety can negatively affect performance [16,43]. Individuals who know they will be judged by experts produce less novel ideas [17]. Many critique providers are aware that public feedback can be emotionally fraught; consequently they take care to temper criticism [53] and supplement critique with praise [28]. Anxiety may increase when people believe their worth as a person is part of what's being assessed [16,33]. For this reason, many educators and parents use language that critiques the *work* and the *behavior*, rather than the *person* [19,43].

Creating multiple designs may help both critique providers and creators separate egos from artifacts. When asked for feedback, people provide more substantive critique when presented with multiple design alternatives [53]. People react less negatively to critique when they create multiple alternatives in parallel, rather than serially [18]. This prior work studied individual behavior; this paper analyzes the social effects.

This paper hypothesizes that sharing multiple designs—rather than one—improves group rapport and increases the rate at which people exchange ideas.

Hypothesis 2: Sharing multiple prototypes leads to more productive dialogue and better group rapport.

This study measured peer interaction by counting speech turns by each partner [40,41]. Also, five questions posed before and after the group discussion assessed individual views of their group's rapport.

Sharing multiple ideas facilitates conceptual blending

When collaborating, groups often merge properties of different concepts [20]. Sometimes, these blends directly inherit properties [26], other times blends spawn new emergent features [22,47]. Blending can be highly structured, as in morphological design [57], but is more commonly ad

hoc. When concepts are dissimilar blending them yields a more ambiguous artifact [56].

In design, conceptual ambiguity can provide a generative resource [23,35]. Sharing multiple designs may help collaborators blend ideas. The process of comparing and contrasting alternatives helps people create higher-level structures [24]; these structures also help collaborators understand and communicate the rationale behind design decisions [38].

This paper hypothesizes that sharing multiple prototypes facilitates conceptual blending and that collaborators will use more surface-level and thematic features from their partner's work.

Hypothesis 3: *Sharing multiple prototypes leads to more effective conceptual blending.*

This study measures conceptual blending by counting features that migrate from one partner's preliminary designs to the other's final design. Independent raters also judged the similarity of partner's designs before and after the pair shared their work.

Finally, this paper hypothesizes that sharing multiple designs leads to better performance due to a confluence of three factors: individuals explore more divergent ideas; groups have stronger dialogue and rapport; and the final design exhibit more effective conceptual blending.

Hypothesis 4: *Sharing multiple prototypes produces better design results.*

This study measures design quality by gathering click-through performance metrics on advertisement designs and by recruiting professionals, clients, and other independent judges to rate ads.

METHOD

A between-subjects study manipulated prototyping practices prior to a group critique. Web advertising was chosen as the design task because it fulfills the following criteria:

- Quality can be measured objectively and subjectively;
- Participants need minimal artistic or engineering ability to either create or critique ads;
- Individuals can complete tasks within a single lab session;
- Solutions demonstrate creative diversity and a range of performance quality.

Study Design

Participants all created Web ads for the same client, *FaceAIDS.org*. The study allocated equal time for individual design and group discussions across three conditions. In the *Share Multiple* condition, participants created three prototype advertisements and shared all three during the group discussion. In the *Share Best* condition, participants created three prototype ads and chose one to share during the group discussion. In the *Share One* condition, participants spent the entire individual design time on a single ad to share during the group discussion.

Participants

We recruited 84 participants through papers fliers, online advertisements, and email lists. Two individuals arrived concurrently to form study pairs. Each pair was assigned to one of three conditions using a stratified randomization approach; the study balanced for gender (41 females) and graphic design knowledge across pairs and conditions. Ten true-or-false questions assessed graphic design knowledge (see Appendix A); participants were deemed *experienced* if they correctly answered eight or more (36 did). Participants who scored below eight were deemed *novices*. Participants' average age was 26.5; three-fourths were students.

Procedure

The experiment comprised the following steps: consent form, icebreaker, tool training, practice ad, design brief, individual prototyping, discussion, final design, group interview, and final debriefing. Questionnaires collected demographic and self-report assessments. The icebreaker, group discussion, and group interview were co-located and video-recorded. All other procedures took place in separate rooms at individual workstations with no video recording. For 120 minutes of participation, subjects received \$20 USD cash.

Icebreaker activities

Partners collaborated on three icebreaker activities for three minutes each. They built a tower with toy blocks, played the game Operation, and generated a list of animal names beginning with 'M' (e.g., monkey).

Graphic design tool training

At separate workstations, partners viewed a five-minute video about the Web-based graphic design tool [1]. Then, using the tool, participants replicated a graphic unrelated to the main task. All participants replicated the graphic in less than ten minutes. None had used the tool before. Selecting a novel tool avoids confounds from participant's tool-specific expertise.

Design brief

A five-minute video described participants' main design activity: to create an ad for *FaceAIDS*, a non-profit organization dedicated to global health equity and social justice [2]. In the video, the organization's executive director outlined four goals: reach out to students interested in starting local chapters of *FaceAIDS*, increase traffic to the *FaceAIDS* Web site, impress three judges from the *FaceAIDS* organization, and create ads with effective graphic design. A paper version of the design brief was available for the group discussion (see Appendix A).

Individual Design Period

All participants had 30 minutes for individual design. In the *Share Multiple* and *Share Best* conditions, participants started a fresh design every 10 minutes. At the end of this period, *Share Best* participants were prompted to select one design to be critiqued by the study partner. After the design period, a study proctor printed ads for the group discussion.

Group discussion

Participants sat together and viewed a print out of their partner's design(s). The study proctor set a timer for five minutes and then instructed the pair: "Examine your peer's design concept(s) and then provide a critique. What advice would you provide? Please speak aloud." After this, the proctor set another five-minute timer and instructed: "Now spend another 5 minutes discussing what you think is the most effective way to satisfy the design brief." After that, participants were instructed to return to their individual workstations to create a final ad design.

Final design period

Participants individually created another advertisement and were instructed that this final ad would be rated by judges and hosted in a live ad campaign.

Group interview

The study concluded with an open-ended group interview. A study proctor used an interview guide and followed up with related questions. These questions provided guidance for the final interview; the exact order and phrasing varied.

- Describe how you arrived at your final design.
- Explain the difference between your two final ads.
- How much did the group discussion affect what you did in your final ad design?
- How did your peer's critique affect your ad design?
- To what extent were you able to reach agreement on the final design concept?

Dependent Measures

Performance

After the experiment, the final graphic ads were hosted on Google AdWords [3] for a 12-day campaign. Design performance was determined through two objective measures:

- Click-through rates (CTR): number of clicks divided by the number of impressions, and
- Google Analytics [4] on the target client Website: total time spent and number of pages visited from each ad.

Ads were also independently judged by 30 individuals: three clients from *FaceAIDS*, six ad professionals, and twenty-one people recruited from Mechanical Turk, an online crowdsourcing system for paying workers for short tasks [5]. This collection of raters provided important—and different—audience perspectives. Each judge read the *FaceAIDS* design brief and viewed ads in random order. For each ad, they estimated (on a 7-point scale) each ad's performance in an online campaign (see Figure 2).

In an online ad campaign, how well will this ad perform?

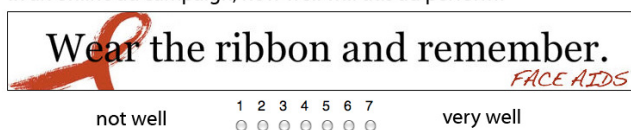


Figure 2 Quality rating: Judges rated (on a 1-7 scale) how well each ad will perform in an online campaign.

How similar are these advertisements?



Figure 3 Similarity rating: Judges viewed a pair of ads and rated their similarity on a seven-point scale. This pair's average rating was 5.7. (The overall average was 3.6.)

Individual design exploration

Exploring a diverse set of ideas can help people examine the space of designs and their relative merits [13]. To obtain a measure of idea diversity, ten independent raters assessed pair-wise similarity of all combinations of individual participant's ads (see Figure 3). Raters recruited from Amazon Mechanical Turk assessed similarity on a scale from 1 to 7 (not similar to very similar).

Change in group rapport

At two points—after the icebreakers and after the discussion—five questions asked individuals to assess their group rapport. The change between these two points measures the discussion's impact on group rapport. Four questions originate from the Subjective Value Inventory (SVI), an assessment of viewpoints on negotiation [15]. The relationship questions from the SVI provide a systematic measure of a group rapport; they assess partners' feelings about the relationship in terms of overall impressions, satisfaction, trust, and foundations for future interaction. The fifth question derives from the Inclusion of Self in Others Scale (see Figure 4), a measure of someone's sense of connectedness with another [7]. The questions asked:

- What kind of overall impression did your peer make on you?
- How satisfied are you with your relationship with your peer as a result of the interaction?
- Did the interaction make you trust your peer?
- Did the interaction build a foundation for future interactions with your peer?
- Please check the picture below which best describes your relationship with your peer:

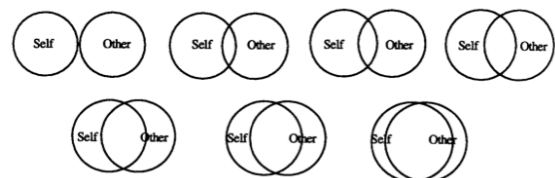


Figure 4 Inclusion of Self in Others Scale: Illustration reprinted from Aron *et al.* [7]

Conversational turn taking

In the group discussion, partners exchanged ideas. A coder recorded the start time and duration of each group member's utterances. This provided the overall number of speech turns by each partner, the total amount of speaking,



Figure 5 Design feature sharing: The two partner ads above have **three** commonalities: images, phrasing, and background color

the ratio of time spent by the more conversationally-dominant partner (high and low talkers), and the frequency of turns per minute of interaction.

Design feature sharing

For each final ad, we counted cross-pollinated features in five categories: word phrases, background color, images, layout, and styles (*i.e.*, fonts, rotations, *etc.*). Cross-pollination was a binary value for each category. A category received a mark if a participant's final ad exhibited a feature that was present in their partner's shared provisional ad, but *not* in their own provisional ad(s) (see Figure 5).

Group consensus

As an aggregate measure of group consensus, independent raters assessed pair-wise similarity between partner ads. The similarity assessment contrasted ads created before and after the discussion. Ten raters recruited from Mechanical Turk assessed similarity on a scale from 1 to 7 (not similar to very similar) (see Figure 3). If the designs are more similar after the discussion, it suggests that partners converge around similar concepts.

RESULTS

Participants created a wide variety of ad designs, demonstrating a range of quality. The highest-rated ads tended to be original, visually appealing, and cleverly touched on themes relevant to *FaceAIDS* (see Figure 6, left). The highest rated ads did not always garner the highest click-through rates (see Figure 6, right); they used images that reflected the client's brand, but exhibit poorer graphic design.

FaceAIDS reviewed the ads before they appeared online. The client found four of the ads to have inappropriate negative imagery, and requested they not be shown. Three of these were from *Share Best*; one was from *Share Multiple*.

In total, the ad campaign generated 239 clicks on 274,539 impressions (ad appearances). The total advertising costs were \$362 USD (an average of \$1.51 per click).

Participants in the *Share Multiple* condition produced higher-quality designs (better click-through rates and higher subjective ratings) and created more diverse designs. Pairs of participants in the *Share Multiple* condition reported a greater increase in rapport, exchanged more verbal information, and shared more ideas. Moreover, ads by more experienced participants received higher rating than novices; the prototypes created by experienced participants were less diverse than novices.

Sharing multiple led to higher quality designs

Ad campaign results

A chi-squared analysis examined ad campaign performance for all 12 days. *Share Multiple* ads had 98,867 impressions with 106 clicks, *Share Best* ads had 77,558 impressions with 57 clicks, and *Share One* ads had 98,038 impressions with 76 clicks (Figure 1 summarizes the average clicks per million impressions). *Share Multiple* ads had a significantly higher click-through rate ($\chi^2 = 4.72$, $p < 0.05$).

An analysis of variances was performed with condition (*Share Multiple*, *Share Best*, and *Share One*) and graphic design score (*experienced* or *novice*) as factors and total time spent and pages visited as dependent variables. Between conditions, there were no differences for total time spent ($F(5,202) = 0.808$, $p > 0.05$) or number of pages visited from each ad ($F(5,202) = 0.461$, $p > 0.05$).

Graphic design experience did not effect campaign results

Ads created by participants who scored high on the graphic design exam garnered 110 clicks on 128,783 impressions; novice ads had 129 clicks on 145,756 impressions. This was not a significant difference ($\chi^2 = 0.08$, $p > 0.05$). Experienced participants benefited more from the manipulation than novices did (see Table 1). Experienced participants in the *Share Multiple* condition outperformed experienced participants in the *Share Best* ($\chi^2 = 3.95$, $p < 0.05$) and *Share One* conditions ($\chi^2 = 8.33$, $p < 0.05$). There were no ad performance differences between conditions for novices.

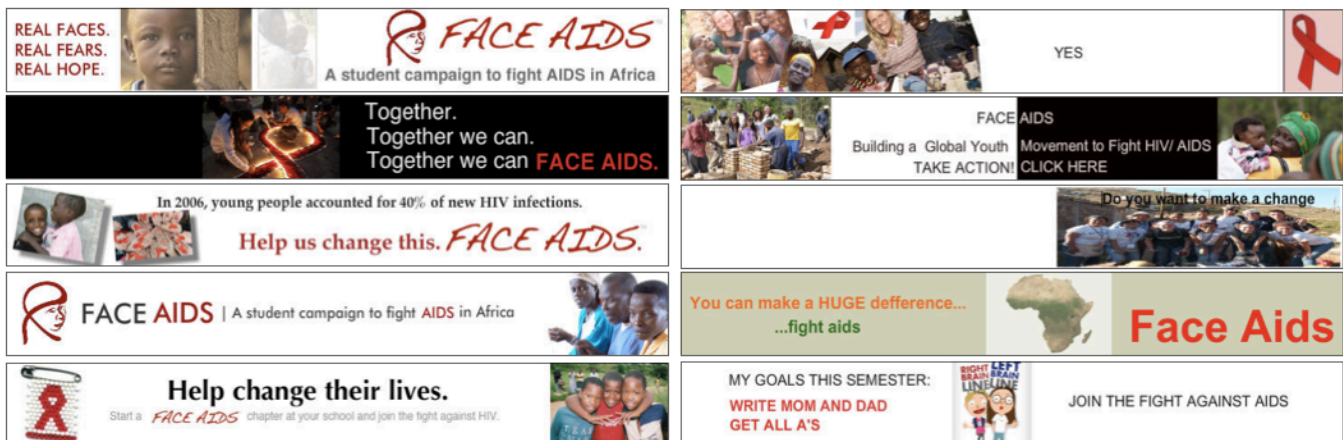


Figure 6 Top five highest-rated ads (left); ads with top five click-through rates (right)

	Experienced	Novice
<i>Share Multiple</i>	1125.3 (818.0)	991.7 (482.2)
<i>Share Best</i>	704.3 (422.2)	758.9 (888.5)
<i>Share One</i>	540.3 (500.0)	905.5 (371.7)

Table 1 Online performance in clicks per million impressions for condition and experience (std dev in parentheses). Experienced participants created better ads and were more affected by condition.

An ANOVA showed that experience did not significantly affect total time spent ($F(5,202)=0.091$, $p>0.05$) or number of pages visited from each ad ($F(5,202)=0.076$, $p>0.05$).

Quality ratings

Thirty raters judged all final ads; their average rating was 3.74 on a 7-point scale. An analysis of variances was performed with condition (*Share Multiple*, *Share Best*, and *Share One*) and graphic design score (*experienced* or *novice*) as factors and performance rating as the dependent variable. The *Share Multiple* condition ($\mu=3.89$, $SD=1.82$) outperformed the other conditions ($F(2,2519)=5.075$, $p<0.05$). The difference between the *Share Best* ($\mu=3.63$, $SD=1.78$) and *Share One* ($\mu=3.71$, $SD=1.71$) conditions was not significant ($t(1678)=0.983$, $p>0.05$).

	Clients	Ad pros	Turkers
<i>Share Multiple</i>	4.06 (1.70)	2.95 (1.63)	4.14 (1.80)
<i>Share Best</i>	3.45 (1.77)	2.76 (1.63)	3.90 (1.74)
<i>Share One</i>	3.79 (1.69)	2.85 (1.55)	3.94 (1.68)

Table 2 Average ratings (std dev in parentheses). All rater types (clients, ad pros, and turkers) rated *Share Multiple* ads higher.

Professionals, clients, and turkers (workers on Amazon Mechanical Turk) all rated *Share Multiple* ads higher than the other conditions (see Table 2). Clients ($\mu=3.77$, $SD=1.73$) and turkers ($\mu=3.99$, $SD=1.74$) ratings were higher on average than those by ad professionals ($\mu=2.85$, $SD=1.60$) ($F(2,2519)=86.961$, $p<0.05$). This differential between advertising professionals and other stakeholders is consistent with prior work [18].

Graphic design experience led to better ratings

Participants who scored highly on the graphic design exam ($\mu=4.10$, $SD=1.709$) significantly outperformed those who scored poorly ($\mu=3.48$, $SD=1.773$), ($F(1,2519)=74.613$, $p<0.05$). The ANOVA shows that novices benefited more from the manipulation than experienced participants did ($F(2,2519)=3.536$, $p<0.05$) (see Table 3). This differential gain is the opposite from the click-through rate, where experienced participants benefited more from sharing multiple designs.

	Experienced	Novice
<i>Share Multiple</i>	4.11 (1.71)	3.68 (1.90)
<i>Share Best</i>	4.19 (1.71)	3.31 (1.74)
<i>Share One</i>	4.01 (1.71)	3.48 (1.67)

Table 3 Average ratings by condition and experience (std dev in parentheses). Experienced created higher-rated ads; novices were more affected by condition.

Sharing multiple led to more individual exploration

Raters from Amazon Mechanical Turk deemed *Share Multiple* ads to be most divergent. An analysis of variances was performed with condition (*Share Multiple*, *Share Best*, and *Share One*) and graphic design score (*experienced* or *novice*) as factors and pair-wise similarity rating as the dependent variable. *Share Best* ($\mu=3.99$, $SD=1.96$) and *Share One* ($\mu=5.45$, $SD=1.86$) ads were deemed significantly more similar than *Share Multiple* ads ($\mu=3.85$, $SD=1.93$) ($F(2,3640)=82.07$, $p<0.05$). T-tests confirmed that *Share Multiple* ads were more diverse than *Share Best* ads ($t(3358)=2.077$, $p<0.05$), and *Share Best* ads were more diverse than *Share One* ads ($t(1958)=11.66$, $p<0.05$).

Experienced participants created ads that were deemed significantly more similar ($\mu=4.20$, $SD=1.96$) than those who scored poorly ($\mu=3.91$, $SD=1.99$) ($F(1,3640)=7.692$, $p<0.05$). There was no interaction effect between condition and prior experience.

Group rapport rose for partners who shared multiple

Group rapport increased in the *Share Multiple* condition ($\mu=0.89$, $SD=3.06$) compared to the other conditions ($F(2,83)=4.147$, $p<0.05$). It was the only condition where participants rated rapport after the discussion higher than before (see Table 4). Overall, group rapport dropped from an average of 24.5 to 23.5 ($t(83)=2.050$, $p<0.05$). Pair-wise t-tests confirmed significant differences between the *Share Multiple* condition and the *Share Best* ($t(54)=2.816$, $p<0.05$) and *Share One* conditions ($t(54)=2.539$, $p<0.05$).

	Before design critique	After design critique	Group rapport shift
<i>Share Multiple</i>	$\mu=24.6$ (4.35)	$\mu=25.5$ (4.66)	+0.89
<i>Share Best</i>	$\mu=24.0$ (5.24)	$\mu=22.3$ (4.79)	-1.75
<i>Share One</i>	$\mu=24.9$ (5.18)	$\mu=22.8$ (5.86)	-2.11

Table 4 Individual views of group rapport rose in the *Share Multiple* condition; it dropped in other conditions (std dev in parentheses)

Share Multiple partners took more conversational turns

An video analysis of speech duration during the group discussion showed that participants in the *Share Multiple* condition had significantly more frequent verbal exchanges (a higher number of speaker turns per minute of speaking time) than other conditions ($F(2,39)=3.506$, $p<0.05$). *Share Multiple* pairs averaged 12.1 ($SD=4.99$) turns per minute, compared to 9.1 ($SD=2.62$) and 8.6 ($SD=2.86$) turns per minute, for *Share Best* and *Share One*, respectively. There were no significant between-condition differences for total number of speaker turns ($F(2,39)=0.695$, $p>0.05$), total speaking time ($F(2,39)=1.057$, $p<0.05$), or the ratio of high and low talkers ($F(2,39)=0.092$, $p<0.05$).

Share Multiple pairs borrowed more features

In total, *Share Multiple* partners borrowed 32 features, *Share Best* 18, and *Share One* 19 (see Table 5). (The theoretical maximum for each condition is 140: 28 participants, 5 categories.) Participants in the *Share Multiple* condition borrowed significantly more features overall ($\chi^2=4.05$, $p<0.05$).

	<i>Share Multiple</i>	<i>Share Best</i>	<i>Share One</i>
Word phrases	15	9	6
Background color	3	0	2
Images	10	6	7
Layout	3	2	3
Surface patterns	1	1	1
Total	32	18	19

Table 5 Participants in the *Share Multiple* condition borrowed more features from their partners than other conditions

***Share Multiple* pairs reached a better consensus**

Independent judges rated the partner ad similarity before and after the discussion. The similarity change provides a measure of shared perspective. Overall, final ads were more similar ($\mu=3.40$, $SD=1.91$) than initial ads ($\mu=2.68$, $SD=1.64$) ($t(3078)=8.107$, $p<0.05$). Similarity increased more for the *Share Multiple* condition (0.91) than the *Share Best* (0.55) or *Share One* conditions (0.52) (see Table 6).

	Before design critique	After design critique	Similarity shift
<i>Share Multiple</i>	2.59 (1.55)	3.50 (1.91)	0.91
<i>Share Best</i>	2.75 (1.71)	3.30 (1.97)	0.55
<i>Share One</i>	2.87 (1.81)	3.39 (1.85)	0.52

Table 6 Pairs designs in the *Share Multiple* condition increased in similarity more than other conditions (st. dev. in parentheses)

DISCUSSION

The results supported all hypotheses. Sharing *multiple* designs led to several kinds of better outcomes. Simply *creating* multiple prototypes was not sufficient. Both the *Share Best* and *Share Multiple* participants created several alternative designs. The benefits were only realized if participants *shared* multiple designs. It's important to remember that participants worked on the same task for the same amount of time. The only variable was how many prototypes they created and shared. Why did this simple act affect performance? This section outlines several reasons for the differential outcomes, illustrating these with interview excerpts.

Does producing and sharing multiple designs impact how people individually explore the conceptual landscape? The *Share Multiple* and *Share Best* conditions explored significantly more broadly than the *Share One* condition. As this study and prior work found [18], rapidly producing alternatives—as opposed to refining one idea—results in more variance. Creative work often benefits from broadly exploring possibilities before choosing a direction to refine [13]. Options must first be on the table to be consciously chosen. By forcing individuals to embody more ideas, both the *Share Multiple* and the *Share Best* strategies support divergent individual exploration. *Share Multiple* goes further by exposing more perspectives. As one *Share Multiple* participant described seeing her partner's designs: “they were completely different from mine and I was like holy hell, that's pretty good. I didn't think about that.” Another par-

ticipant claimed that “getting a different perspective...helped and also seeing different ideas - not flaws in mine, but different ideas in his that I'd like to borrow.” Discussing multiple designs encouraged the construction of conceptual structures to compare and contrast within and across participants.

The study found that experienced participants created less diverse designs than novices; their ads were also rated higher. As Schön describes, expert designers can rapidly construct entailments, mentally simulating design “moves” and their consequences [44]. Experienced designers ignore or disregard the obviously bad options. This foresight allows experts to sample a narrower, yet more promising part of the conceptual landscape.

Does sharing multiple prototypes improve design collaboration? Across all conditions, participants reported an overall decline in group rapport from before to after the discussion. This is not surprising given the sensitive nature of critique. As one *Share Best* participant admitted, “she didn't make me feel terrible about what I produced... even though I...I... she didn't make me feel like a total failure.” Having only selected her “best” prototype, this individual's sense of risk may have disrupted group interaction. The notion of taking personally a partner's critique appears obviated to some extent when participants distribute effort across multiple designs. Reports of group rapport actually increased when participants shared multiple designs. Similarly, studies of group design show that when teams generate lots of ideas, people feel more shared ownership and stronger team cohesiveness [25,50].

Pairs in the *Share Multiple* condition had significantly more frequent verbal exchanges than the other conditions. As one *Share Multiple* participants said, “being able to see the other person's designs and actually bounce ideas back and forth... that helped clarify what was good design and what wasn't.” More frequent exchanges enabled participants to delve into design tradeoffs and the context for advertising, as one participant said, “it got me thinking about who would click on an ad and why someone would click on an ad.” Such statements illustrate some partners' willingness to discuss more fundamental changes to their solutions.

Does sharing multiple prototypes increase conceptual blending? Compared to conditions where participants shared only one ad design, participants in the *Share Multiple* condition integrated more features and modified their designs to be more like their partner's. Participants often talked about the process of merging designs. One *Share Multiple* participant said, “we agreed we like elements of mine and I really like some elements of one of his and we just kind of did a mash-up and combined them.” In contrast, a *Share Best* participant said, “we thought about some ideas, but we didn't really get to a consensus of what we were going to design.” Likewise, a *Share One* participant said, “I didn't really get a lot of things to change on mine, so I just stuck with what I had.” This notion of “sticking” with an idea did not surface in the *Share Multiple* condition.

Direct comparison can help people understand underlying principles [24]. Since *Share Multiple* participants viewed several of their partners' designs, they could utilize comparison processes to see deeper patterns. One participant in the *Share Multiple* condition saw beyond surface-level similarities, and said "(our) ads look different, but I feel like in general it's the same message that's getting across." This pair claimed to be connecting their ideas on a more thematic level. Forming a stronger understanding of their partner's design rationale may be one reason participants came closer to consensus and produced better results.

Reflections on Method

The advertising paradigm described in this paper helps provide experimenters leverage when studying creative work. It offers strong quantitative benchmarks through its Web analytics, captures the views of many stakeholders, and provides measures of several different types of outcomes. While using an advertising platform to measure analytics has clear benefits in terms of ecological validity, the goals of an ad platform can sometimes be in conflict with scientific goals. For others interested in this approach, we share three challenges.

First, many advertisers show ads differentially. If an ad (or cohort of ads) performs well, it is shown more. If initial performance is poor, it is shown less. Experimenters must take care to insure that all ads are shown sufficiently, and that interactions between impressions and click-through rate are minimized. There's a reason that some people make a living managing online ad campaigns.

Second, at first blush, it can appear that advertising pipelines have an unlimited capacity for showing ads, and consequently scale effortlessly. This is not the case. (Except, perhaps, for Justin Bieber.) Even with a major client and the world's largest advertising network, eighty ads pushes the limits of how many alternate ads can be simultaneously shown. Showing more ads would prevent each one from receiving enough impressions for statistical analysis. Even in the age of the Web, experimenters must be judicious.

Third, with many measures, which one is "right"? Raters assess aesthetic quality; clicks assess bottom line. (At least in the short term.) In our study, online click performance did not correlate with overall rating ($R^2 = 0.018$, $F(1,79) = 2.445$, $p > 0.05$, $b = 0.174$). Some unattractive ads receive many clicks; the Web has a preponderance of such examples. Do clicks or positive reviews by different stakeholders define the success of an advertisement? As in brainstorming, design has several objective functions [50].

CONCLUSIONS AND FUTURE WORK

This paper found that when people produce and share multiple alternatives at group discussions, they explore more diverse ideas, integrate more of their partner's features, engage in more productive design conversations, and ultimately, create higher-quality work. Many designers already practice this approach. This research seeks to develop a theoretical understanding of creative work to help practitioners and students solve design problems more effectively.

These results suggest that many more practitioners and teachers might beneficially adopt a "share multiple" strategy. More broadly, this work raises several important questions.

First, (how) do these results generalize to different types of groups? In this study, participants were independently recruited, with no prior collaboration, and had comparable authority. In most professional work, collaboration is longitudinal, and power relationships and social dynamics are more complex.

Second, how might groups be alternatively structured? For example, what if dyads performed complementary tasks rather than equivalent ones? Such work could explore strategies for cross-functional teams. One example of a cross-functional strategy is jigsaw learning, where different students are responsible for complementary parts of a topic [8]. Jigsaw learning has been found to increase learning outcomes, especially in heterogeneous classrooms.

Third, recent research on "the crowd within" suggests that at least some of the benefits of aggregating many people's work can be accomplished by providing individuals with a structured approach to considering alternatives [54]. This study witnessed several benefits of group discussion; can some of these be achieved through structured reflection?

Fourth, this study's three conditions were: create one and share one, create three and share one, create three and share three. This leaves an unexplored cell: create one design and share three designs. How might designers share multiple ideas without creating them? One strategy would be to have designers supplement their own creations with previously created examples. An alternative would be for software to (semi-)automatically synthesize design alternatives [27,34].

Fifth, the benefits of rapidly creating and sharing multiple alternatives are myriad. How might software tools help designers work explore more broadly? Initial results are promising [27,34]; more exciting work remains.

REFERENCES

1. Web graphic editing tool. <http://www.flashimageeditor.com>.
2. FaceAIDS Organization. <http://faceaids.org>.
3. Google Ad Words. <https://adwords.google.com>.
4. Google Analytics. <http://www.google.com/analytics>.
5. Amazon Mechanical Turk. <https://www.mturk.com/mturk>.
6. Arkes, H.R. and Blumer, C. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes* 35, 1 (1985), 124-140.
7. Aron, A., Aron, E.N., and Smollan, D. Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596-612.
8. Aronson, E., Bridgeman, D., and Geffner, R. Interdependent Interactions and Prosocial Behavior. *Journal of Research and Development in Education* 12, 1 (1978),

- 16-27.
9. 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.
10. Bao, P., Gerber, E., Gergle, D., and Hoffman, D. Momentum: getting and staying on topic during a brainstorm. *Proc of conf on Human factors in computing systems*, ACM (2010), 1233-1236.
11. Brandt, J., Dontcheva, M., Weskamp, M., and Klemmer, S.R. Example-centric programming: integrating web search into the development environment. *Proc of conf on Human factors in computing systems*, ACM (2010), 513-522.
12. Brereton, M., Cannon, M., Mabogunje, A., and Leifer, L. Collaboration in Design Teams: How Social Interaction Shapes the Product. In *Analyzing Design Activity*. Wiley, 1996.
13. Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, 2007.
14. Cross, N. Expertise in design: an overview. *Design Studies* 25, 5 (2004), 427-441.
15. Curhan, J.R., Elfenbein, H.A., and Xu, H. What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation. *Journal of Personality and Social Psychology* 91, 3 (2006), 493-512.
16. Dannels, D.P. and Martin, K.N. Critiquing Critiques: A Genre Analysis of Feedback Across Novice to Expert Design Studios. *Journal of Business and Technical Communication* 22, 2 (2008), 135-159.
17. Diehl, M. and Stroebe, W. Productivity Loss In Brainstorming Groups: Toward the Solution of a Riddle. *Journal of Personality and Social Psychology* 53, 3 (1987), 497-509.
18. Dow, S., Glassco, A., Kass, J., Schwarz, M., Schwartz, D.L., and Klemmer, S.R. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy. *Transactions on Computer-Human Interaction* 4, (2010).
19. Dweck, C. *Mindset: The New Psychology of Success*. Ballantine Books, 2007.
20. Fauconnier, G. and Turner, M. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, 2003.
21. Felps, W., Mitchell, T., and Byington, E. How, When, and Why Bad Apples Spoil the Barrel: Negative Group Members and Dysfunctional Groups. *Research in Organizational Behavior* 27, (2006), 175-222.
22. Finke, R.A., Ward, T.B., and Smith, S.M. *Creative Cognition: Theory, Research, and Applications*. The MIT Press, 1996.
23. Gaver, W.W., Beaver, J., and Benford, S. Ambiguity as a resource for design. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2003), 233-240.
24. Gentner, D., Loewenstein, J., and Thompson, L. Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology* 95, (2003), 408, 393.
25. Gerber, E. Prototyping Practice in Context: The Psychological Experience in a High Tech Firm. *Journal of Design Studies*, (2010).
26. Hampton, J.A. Inheritance of attributes in natural concept conjunctions. *Memory & Cognition* 15, 1 (1987), 55-71.
27. Hartmann, B., Yu, L., Allison, A., Yang, Y., and Klemmer, S.R. Design as exploration: creating interface alternatives through parallel authoring and runtime tuning. *Proc of the conf on User interface software and technology*, ACM (2008), 91-100.
28. Hyland, F. and Hyland, K. Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing* 10, 3 (2001), 185-212.
29. Iyengar, S.S. and Lepper, M.R. When choice is demotivating: can one desire too much of a good thing? *Journal of Personality and Social Psychology* 79, 6 (2000), 995-1006.
30. Janis, I.L. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Wadsworth Publishing, 1982.
31. Jansson, D. and Smith, S. Design Fixation. *Design Studies* 12, 1 (1991), 3-11.
32. 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.
33. Kosara, R. Visualization Criticism - The Missing Link Between Information Visualization and Art. *Proc of the Conf on Information Visualization*, IEEE Computer Society (2007), 631-636.
34. Lee, B., Srivastava, S., Kumar, R., Brafman, R., and Klemmer, S.R. Designing with interactive example galleries. *Proc of the conf on Human factors in computing systems*, ACM (2010), 2257-2266.
35. Leifer, L. Dancing with Ambiguity: design thinking in theory and practice. 2010.
36. Mark, G., Gonzalez, V.M., and Harris, J. No task left behind?: examining the nature of fragmented work. *Proc of the conf on Human factors in computing systems*, (2005), 321-330.
37. Marsh, R.L., Landau, J.D., and Hicks, J.L. How examples may (and may not) constrain creativity. *Memory & Cognition* 24, 5 (1996), 669-680.
38. Moran, T.P. and Carroll, J.M. *Design Rationale: Concepts, Techniques, and Use*. CRC Press, 1996.
39. Nickerson, R.S. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2, (1998), 175--220.
40. Pentland, A. Learning Communities — Understanding Information Flow in Human Networks. *BT Technology Journal* 22, 4 (2004), 62-70.
41. Ranganath, R., Jurafsky, D., and McFarland, D. It's not

you, it's me: detecting flirting and its misperception in speed-dates. *Proc of Conf on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics (2009), 334-342.

42. S. Smith, T.B. Ward, and J.S. Schumacher. Constraining effects of examples in a creative generation task. *Memory & Cognition* 21, (1993), 837-845.
43. Schon, D.A. *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. Jossey-Bass, 1990.
44. Schon, D.A. *The Reflective Practitioner: How Professionals Think in Action*. Ashgate Publishing, 1995.
45. Schrage, M. *Serious Play: How the World's Best Companies Simulate to Innovate*. Harvard Business School Press, 1999.
46. Schwartz, B. *The Paradox of Choice: Why More Is Less*. Ecco, 2004.
47. Schwartz, D.L. The Emergence of Abstract Representations in Dyad Problem Solving. *Journal of the Learning Sciences* 4, 3 (1995), 321.
48. Stroebe, W. and Diehl, M. Why Groups are less Effective than their Members: On Productivity Losses in Idea-generating Groups. *European Review of Social Psychology* 5, (1994), 271.
49. Sutton, R. Eight Tips for Better Brainstorming. *BusinessWeek: Innovation*, 2006.
50. Sutton, R. and Hargadon, A. Brainstorming groups in context: effectiveness in a product design firm. *Administrative Science Quarterly*, (1996).
51. Taylor, D., Berry, P., and Block, C. Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking? *Administrative Science Quarterly* 3, 1 (1958), 23-47.
52. Thomke, S. and Nimgade, A. IDEO Product Development. *Harvard Business School Case*, (2000).
53. 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.
54. Vul, E. and Pashler, H. Measuring the crowd within: probabilistic representations within individuals. *PSYCHOLOGICAL SCIENCE*, (2008), 645--647.
55. Warr, A. and O'Neill, E. Understanding design as a social creative process. *Proc of the conf on Creativity & Cognition*, ACM (2005), 118-127.
56. Wisniewski, E. and Gentner, D. On the combinatorial semantics of noun pairs: {Minor} and major adjustments to meaning. In *Understanding word and sentence*. Amsterdam: North Holland, 1991, 241-284.
57. Zwicky, F. *Discovery, Invention, Research Through the Morphological Approach*. MACMILLAN, 1969.

APPENDIX A: Graphic Design Assessment

Instructions: For each of the statements below, indicate (True or False) whether or not the statement is a rule of graphic design.

- | | | |
|----|--|---|
| 1 | Mix serif and sans serif fonts in order to give variety to the ad. | F |
| 2 | To help balance the ad, leave slightly more space at the top relative to the bottom of the ad. | F |
| 3 | Create a visual separation between the text and the background. | T |
| 4 | Angle the text in order to contrast different parts of the ad. | F |
| 5 | Keep all elements in the ad aligned to one side. | F |
| 6 | Create multiple visual focal points in order to attract attention to the ad as a whole. | F |
| 7 | Use borders or white around text and images to help frame the content. | T |
| 8 | You may use repetition to create a consistent and balanced look. | T |
| 9 | You may break alignment to draw the viewer's attention to important elements in the ad. | T |
| 10 | Draw the viewer's attention to important elements by contrasting scale. | T |

APPENDIX B: Advertising Design Brief

Assignment

You have been hired to design a graphic advertisement for FACEAIDS.org. You will learn to use a new graphic design tool, design provisional ads, and create a final ad to be posted through the Google ad network.

Goals

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

- a) Increase traffic to the *FaceAIDS* website:
<http://www.faceaids.org/>
- b) Reach out to the target audience: students interested in improving global healthcare equality and making a difference in the AIDS epidemic in Africa.
- c) Impress the clients from *FaceAIDS*, who will rate your ads. The client wants an ad that fits their overall aesthetic and theme (see below).
- d) Create ads with effective graphic design. Ad professionals will rate your ads.

What is FaceAIDS?

FaceAIDS is a nonprofit organization dedicated to mobilizing and inspiring students to fight AIDS in Africa. *FaceAIDS* aims to build a broad-based movement of students seeking to increase global health equality. The organization raises awareness and funds, with the goal of increasing global health equality starting with the AIDS epidemic in Africa.

Theme and Aesthetic for the FaceAids Ad

FaceAIDS would like an advertisement that embodies the theme and general aesthetic of the organization. In particular, they are looking to encourage high school and college students interested in getting involved in service or social justice work to start *FaceAIDS* chapters on their campuses, as a leadership development opportunity and a way to join a vibrant, impactful community of like-minded, driven peers. In general they are looking for an ad that is tasteful, creative, professional, visually appealing, and conveys a clear message about the organization.

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.