

How design practices affect results

By Steven P. Dow

In their 2001 book *Art & Fear: Observations On the Perils (and Rewards) of Artmaking*, David Bayles and Ted Orland share a story about a ceramics teacher who divided his class into two groups. He told one half they would be graded on quantity, so they should “produce as many ceramics you can in one quarter, that will be your grade,” while he told the other half they would be graded based on one good ceramic.

Bayles and Orland found that “while the quantity group was busily churning out piles of work—and learning from their mistakes—the quality group had sat theorizing about perfection, and in the end had little more to show for their efforts than grandiose theories and a pile of dead clay.” Iterative deliberate practice led to better results.

Iteration, or, in simpler terms, repetition, is a basic tenet of design practice. (In mathematics and computer science, iteration describes the act of solving a problem by computing a series of approximations, each building on the previous one, to achieve an accurate result.)

While the story about the ceramics class resonates with some people, others might say, “This isn’t how it works in industry where we have real time constraints. Yes, it would be wonderful to try lots of alternatives, but we simply don’t have time.” This raises a question about how design practices affect results—under time constraints, should people iterate or should they focus on refinement?

My research centers on questions about creativity and collaboration. I seek to uncover the cognitive and social factors that affect the often-messy process of design and scientific inquiry, and to investigate these issues within the modern landscape of social media, online gaming and crowdsourcing. Ultimately, I’m interested in understanding and improving how people can design better products, services and systems.

I developed these interests as a post-doc at Stanford University. Stanford has a school of design known as the “dSchool.” About five years ago, they began to teach a problem-solving process known as “design thinking.” Design thinking is an interdisciplinary method of problem solving which puts a premium on prototyping and developing empathy for the target users.

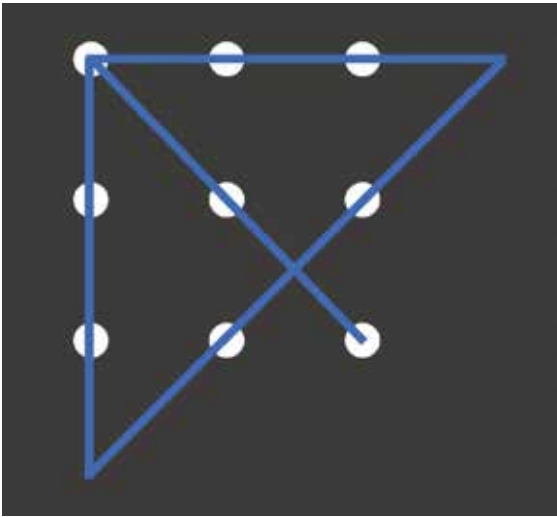
These days, when you step into the dSchool, you’ll find posters urging students to defer judgment, go for quantity, encourage “wild” ideas, build on the ideas of others, have one conversation at a time, stay focused on their topic and think “visually.” Rolling whiteboards, adjustable furniture and ongoing student projects are on display, encouraging collaboration and a free exchange of ideas. These are also signs of commitment to design thinking that say, “Believe in Process.” Often the commitment to particular problem solving strategies rests largely on faith, not on concrete, empirical evidence.

That’s where my research comes in. My colleagues and I have run a series of experiments that examine how prototyping practices affect design results.



| Figure 1

But how can we measure creative thinking skills? Scientists have long been interested in creativity. One classic insight experiment is the nine-dot problem, invented by American psychologist Norman Maier in 1930. (Figure 1) Participants must connect nine dots with four straight lines without ever lifting the pen. The insight that participants often miss is that the lines must extend “outside the box.” (Yes, that’s where the phrase “outside-the-box thinking” originated.) The length of time it takes people to solve the problem provides a dependent measure. (Figure 2)



| Figure 2

When we thought about how design and engineering unfold in real practice, we realized we needed a better Petri dish. Unlike the nine-dot problem, we wanted participants to demonstrate creativity. Real-world design problems have many possible solutions and many different paths to success. We also wanted to have objective and subjective criteria—we needed a good way to contrast solutions people come up with.



| Figure 3

Our insight here was to leverage the “egg drop design” task where participants design a vessel (like the one in Figure 3) to protect a raw egg from a fall. Our dependent variable: how high can you drop the egg without the egg breaking?

In one experiment, in the spirit of the ceramics story, we explicitly examined iteration. Half of our participants were encouraged to rapidly generate new ideas for egg-drop containers; the other half focused on perfecting one design. Everyone came up with very different ideas, with varying degrees of success. (Figure 4)

The results showed quantitatively that, even under tight time constraints, people who were forced to rapidly iterate outperformed those who didn't.



| Figure 4

The experimental results were not particularly surprising, and confirmed the intuitions of the ceramics teacher. Most of us would expect that rapid iteration would yield benefits. But, what was really interesting is what we saw qualitatively in the participant interviews: Independent of study condition, participants tended to pick one idea and stick with it. Participants said things such as, “For some reason, (this design) seems to be the only (way). There needs to be a platform and then as good of a cushion as possible. I don't see any other way.” Or, “I kind of just had one idea and I was going to try to make it work.” Or, “I went with the whole parachute idea ... I had from the beginning.”

Time constraints certainly contributed to participants' limited exploration, but more interestingly, people felt they had fully explored the concepts, and they could not see any other alternatives for the materials.

In design, people often fixate on one solution without considering others. Participants in our egg-drop experiments exhibited a psychological effect known as functional fixation, first studied by German-American psychologist Karl Duncker back in the 1930s. He did a series of experiments that have come to be known as the "Candle Problem." Duncker presented his participants with a table pushed up against a wall. The table held a candle, a box of thumbtacks and a pack of matches. Duncker then asked his participants to affix the candle to the wall so that the wax does not drip on the table. This is a challenging puzzle for most people.

The hidden insight is that the box of tacks—once emptied—can be tacked to the wall and used to hold the candle and catch the dripping wax. Participants in Duncker's experiments often exhibited functional fixation; they viewed the box's only function as a container for tacks. As in our egg drop experiment, once the participants developed an initial idea, they became fixated on making that idea work, instead of exploring different ideas.

Subsequent tests of Duncker's candle problem have showed that if the exact same materials are provided, but the tacks are left outside the box, loose on the table, people are much more likely to solve the puzzle.

Following our egg-drop experiment, we wondered, instead of just iterating solutions to a problem and soliciting feedback on each iteration, what if people created and tried different designs in parallel? (Figure 5)

To answer this empirical question, we gave participants a design task where the solutions were both creatively diverse and objectively measurable. This time, instead of egg-drop vessels, we had participants design Web advertisements for Stanford's *Ambidextrous* magazine, a student-run journal of design and engineering. We were then able to place the ads online and collect objective outcome metrics—how many clicks an ad receives, compared to how many times it's shown.

In the study, each participant created five prototypes and a final design within the same overall time period. In the Serial condition, participants received critiques on one prototype at a time. Participants in the Parallel condition

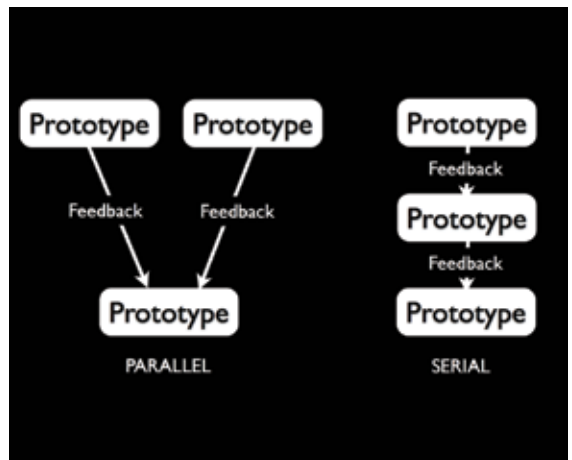


Figure 5

created three prototypes, received critiques on all three, then made two more prototypes, and received critiques again. The critique statements were technical in nature, intended to provide high-level direction, without using explicitly positive or negative language. Importantly, the only difference between conditions was the timing of the critique.

In the end, we got lots of ads. (Figure 6) We took all 33 final participant ads and launched a 15-day ad campaign online. In total, we generated about 1 million ad appearances.

What were the results of this experiment? According to the ad campaign data, Web users clicked more Parallel ads per appearance than Serial ads. Not only did Parallel ads generate more visitors to the *Ambidextrous* website, those visitors spent more time on the client site, so the Parallel ads did better at reaching the target audience.

Moreover, independent expert raters—both ad professionals and the magazine editors—judged the Parallel ads to be better than Serial ads. By all measures, the "Parallel" process outperformed the serial process.

Why did we see this difference? Why does a parallel approach lead to better results?

In design, people often fixate on one solution without considering others. Participants in our egg-drop experiments exhibited a psychological effect known as functional fixation, first studied by German-American psychologist Karl Duncker back in the 1930s.

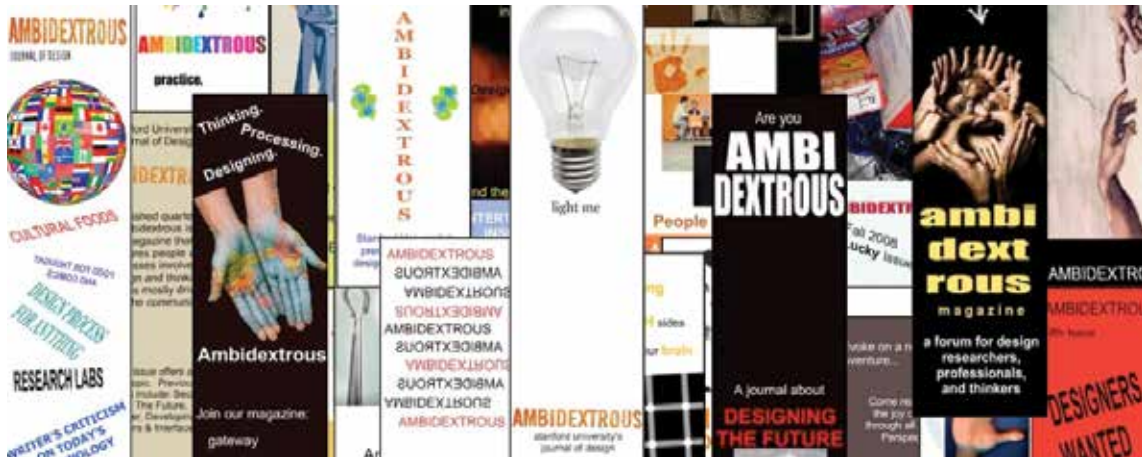


Figure 6

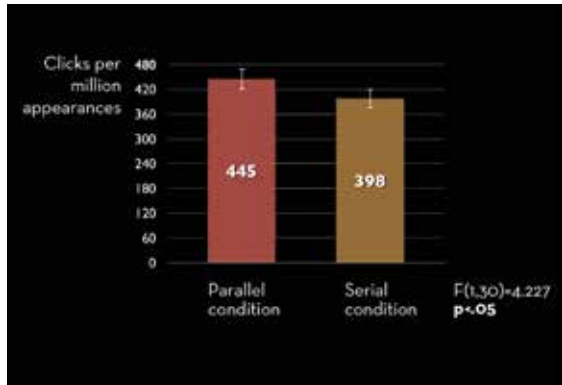
One reason has to do with our basic human ability to draw contrasts. In a 2003 study, Dedre Gentner, Jeffrey Loewenstein and Leigh Thompson compared a traditional case-based learning approach—where participants independently read and described separate case studies—to a more comparative approach. In the comparative approach, participants were explicitly prompted to describe the parallels of both solutions. They found that when prompted to explicitly draw a comparison, participants were nearly three times more likely to understand the principle behind the cases and to transfer what they learned. People do a better job of capturing the underlying structure when they compare.

Going back to our ad study, then, we can surmise that comparing critique statements on two ads side-by-side helped participants extract important graphic design principles.

The interviews provided additional context. In design, feedback is often a double-edged sword: it helps people learn, but it can also damage the ego because people tend to invest emotionally in the things they produce. Eight out of 17 participants in the Serial group reacted negatively, calling the feedback “negative.” One participant told us, “There was a short period (after each critique) where the emotional response overwhelmed any positive logical impact.”

None of the Parallel participants described the critiques as negative. Although the language in the Serial critiques was not any more negative than the Parallel critiques, it was just perceived that way. The groups who worked on ads using the Parallel approach did not emotionally invest in individual solutions; instead, they were open to multiple possible outcomes.

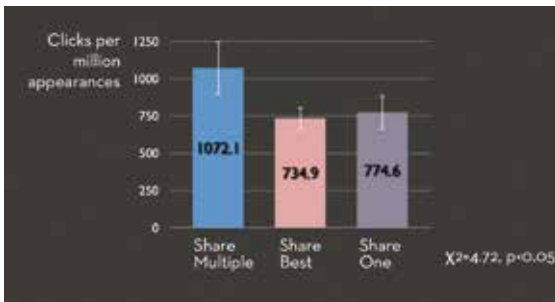
We then asked ourselves, “If people react this way to critiques from a random expert, how would this play out in small groups? In groups, would creating and sharing multiple designs improve the outcome, over just bringing one design?” →



We concluded that better design was a function of better comparison, more individual exploration, more feature-sharing, increased in-group rapport and more conversational turns.

Turing Award winner Fred Brooks once said, “Prototypes can be more articulate than people.” Prototypes help ground communication and embody the entailments of design concepts. However, the presence of a concrete prototype may—for better or worse—focus a discussion on refining that idea, rather than thinking more broadly. Moreover, people tend to polish prototypes to look good in front of their colleagues.

Alternatively, designers may choose to share multiple prototypes at group meetings. In theory, this should help reduce fixation and give group members license to be more candid and critical of their own and other’s ideas. But generating multiple alternatives can also have adverse



| Figure 7

effects. It leaves less time to polish each idea, and increasing the number of options on the table—and the number of implications that arise from these alternatives—may complicate the decision process and jeopardize a group’s ability to achieve consensus.

We hypothesized that sharing multiple prototypes would lead to better results because people would explore more concepts and be more open to adopting and merging new ideas. Again, we had participants design Web advertisements, this time for FaceAIDS.org, a non-profit organization dedicated to fighting AIDS in Africa. Again, we could launch the ads online through Google’s ad networks and collect relative performance metrics.

We recruited 84 participants, balanced for prior experience and gender, and placed them into one of three conditions. In the “share multiple” condition, participants created three ad designs and shared all three in a group meeting with their partner, where they critiqued each other’s ideas. In the “share best” condition, participants created three ads, but then chose only one to share with their partner. In the “share one” condition, participants spent an equivalent amount of time on just one ad, and then shared that with their partner. We chose these three conditions to separate the effects of producing multiple designs from sharing multiple designs.

In all conditions, after the group meeting, each of our 84 participants went back and individually created a final ad. Some were great, some were cliché and some were very clever. These final ads were launched simultaneously in a 12-day ad campaign through Google AdWords. In total, we generated 474,539 impressions. We also had a range of experts—including the FaceAIDS clients—rate the ads on their effectiveness.

The results showed that Web users clicked more ads per appearance created by the “Share Multiple” conditions than either of the other conditions. (Figure 7) Moreover, ad experts and the clients all rated Share Multiple ads higher than the other conditions.

Again we must ask, “Why does creating and sharing multiple designs lead to better results?” Our analysis examined the literature on examples, design exploration, conceptual blending and group rapport.

For one thing, it helps to have more ideas on the table. In a 2010 study, Brian Lee, Scott Klemmer and colleagues of Stanford University’s HCI Group found that people produce better Web designs when given a gallery of examples. Examples expose people to more design features and diverse perspectives.

Were there differences in how participants in each condition explored concepts? How much did that group meeting affect the final designs? Interviews revealed that

participants in the Share Best and Share One conditions tended to stick with what they had. The Share Multiple groups took the best of multiple concepts and blended their ideas.

We quantified this notion of blending concepts by counting features that migrated from one participant's early designs into their partner's final design—similar kinds of images, shared phrasing and a reddish background color. It turns out that participants in the Share Multiple conditions borrowed far more features than pairs in the other conditions. (Figure 8)

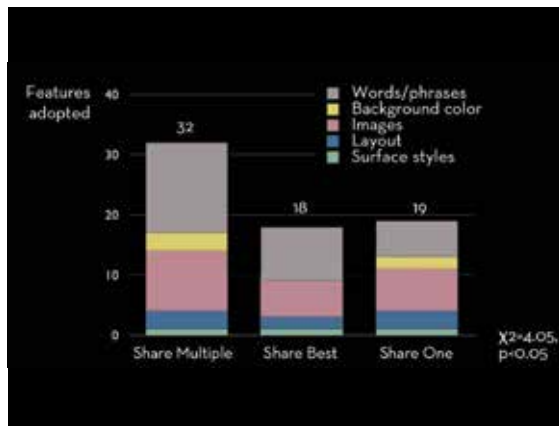
One reason for this was the visibility of work. Much like a design studio, people learn by simply being able to see their peers' ideas. It exposes people to the space of possibilities. Further, during “crits,” Share Multiple participants were

design was a function of better comparison, more individual exploration, more feature-sharing, increased in-group rapport and more conversational turns. A simple change in the process not only produced better designs, it led to more idea sharing and better overall collaboration.

This research on creative process directly informs my future projects. I'm interested in how these phenomena play out in new contexts, particularly with online crowds. I have two new projects, supported by grants from the National Science Foundation, which will examine the intersection of design and crowdsourcing. The first is a collaborative effort with Bjoern Hartmann of the University of California at Berkeley that explores how group dynamics affect creative work done by online crowds. The second is a collaborative effort with Liz Gerber of Northwestern University that examines how we can bring crowdsourcing resources—such as social media, Amazon Mechanical Turk and crowd-funding—into the classroom to help inform student innovations.

Crowdsourcing techniques and web analytics provide an opportunity to do experimental research on creativity with objective outcomes. We have been able to get leverage by giving participants tasks—like Web banner ad design—where the solutions are both creatively different and objectively measurable. Our work shows that design thinking methods have measurable value in the online world, and that simple process changes can lead to better solutions. → □

—Steven Dow is an assistant professor of human-computer interaction at Carnegie Mellon University, where he researches human-computer interaction, creative problem solving, prototyping practices and crowdsourcing. He is a recipient of Stanford University's Postdoctoral Research Award and co-recipient of a Hasso Plattner Design Thinking Research Grant. Dow earned both his M.S. and Ph.D. in human-centered computing at the Georgia Institute of Technology, and a B.S. in industrial engineering at the University of Iowa.



| Figure 8

less invested in a single outcome, so they did a great job of exchanging ideas, while Share One and Share Best participants tended to bottle up and sit there in silence for fear of offending their partner.

Our studies revealed a number of quantitative differences of sharing multiple designs. We concluded that better