A/B Testing: The Most Powerful Way to Turn Clicks into Customers
By Dan Siroker, Pete Koomen and Cara Harshman

CHAPTER

10

Iterate, Iterate, Iterate

The Art of Asking Many Small Questions Rather than One Big One

When test results first start to come in, it's important to recognize that they're likely going to generate more questions than answers. These questions will likely also point to more assumptions you find you want to test. That's okay; in fact, that's how it's *supposed* to work. And that's the perfect starting point for your tests to come.

One of the biggest questions that we had when starting Optimizely was whether people would continue experimenting after they found a "local maximum," that is, the best set of tweaks for their current design or current funnel. We worried that people would spend a few months optimizing a site—a better headline here, different image there—and be done with it. We didn't want businesses to think of A/B testing as a finite, onetime process, and we acknowledged the disconcerting possibility that businesses would feel satisfied if the site worked better than before, and might throw in the towel and declare, "Well, enough of that! We're all done A/B testing."

Much to our delight, we've found that the opposite has occurred. After a few big wins, most people realize that they're just beginning the testing journey—and the success they've enjoyed propels them to keep going.

In month one, you run a couple of tests; you get some wins and you're adding those wins into the production code. Those wins might have been scattered around different pages of your site, since you've likely gone after the low-hanging fruit on various sections of the site. Hopping around like this is great for quick wins that prove value early on. However, you want to

spend the next couple of months concentrating your optimization efforts on a specific area or areas of your site. This approach leads to richer, more valuable test results. It's absolutely fine to use the first month to test a product detail page one day, a landing page another day, and the "About" page the next. But over the next couple months, you want to pick *one* of those pages and focus a series of tests on it. Testing in a series, or *iterative testing*, is the goal here.

Multivariate Testing and Iterative Testing

It's possible to create large-scale tests to assess a number of different variables simultaneously, and these big, compound experiments are known as *multivariate tests*. For instance, two different button colors, three different calls to action, and five different images could all be tested at once, making for a total of $2 \times 3 \times 5 = 30$ different page combinations being shown to different users!

You may recall the example that began this book, from the Obama 2008 campaign: the campaign team used a multivariate test to optimize the best button and best media simultaneously, testing every combination of buttons and media.

One of the questions that we hear frequently is, "When should we use multivariate tests, and when should we use a sequence of single-variable tests?" It turns out this is actually a very nuanced question.

Among the most important things that multivariate tests enable you to discover are *interaction effects* between the different variables or elements you're testing. It might be possible, for instance, that image *X* performs worse than your control, and so

does button *Y*—yet when *X* and *Y* are shown *together* they are significantly superior to the control.

It's a fear of these interaction effects that causes many people initially to assume that *everything* they want to test should be run as one massive multivariate experiment. We think that these fears are often misguided. Here's why.

For one thing, interaction effects do exist, but in *practice* they're relatively rare. Note that the button that performed best in the Obama 2008 multivariate test was also the button that performed best overall *without* taking the media into account. Likewise, the media that performed the best in the multivariate experiment was also the media that performed best overall without taking the button into account. The big multivariate test confirmed that these two indeed worked well together: there weren't any interaction effects that might have complicated those single-variable results.

Even if interaction effects are rare, why not just use multivariate tests for everything, just to be sure? The fact is that multivariate tests require *much more traffic* to produce statistically significant results for each combination. By multiplying the number of permutations, you're also multiplying the number of users that will need to go through your experiment, and the time you'll need to let the experiment run before you get your answer.

Having worked with thousands of customers in a wide variety of industries, we've found that the people who use website optimization most effectively run a series of simple A/B tests, and then incorporate the winner as they go along. They run four or five variations, figure out what works, incorporate that element, and then move on to the next test. When you judge the risk of interaction effects to be low, we strongly recommend that you

test nimbly: start with simple independent A/B tests and *iterate* instead of trying to sort out everything all at once in a multivariate experiment.

"[Testing] gives us humility," says Principal Engineer Dan McKinley at Etsy. "It changes the way we build things." He explains, "There has been a dramatic change in the way we try to build products between 2007 and now, in that now . . . we are going to get there through smaller releases of things that we measure. We are not going to try to do 11 things at once. We'll if possible do 11 things in sequence."

There Are No Universal Truths: Always Be Testing

The world is a very big place; different websites and different products appeal to different people. One of the reasons why A/B testing is so important is that there are *no universal truths* when it comes to design and user experience. If universal truths existed, then A/B testing wouldn't: you'd just look at the rulebook. But because no two audiences are the same—and people are coming from an array of places and perspectives—it's crucial to understand and optimize the experimentation process.

More and more businesses have done so. It's typical to see companies running a number of tests at any time, relative to the time of year, product launches or various campaigns.

What does a long-term A/B testing strategy look like? As Senior Product Marketing Manager Jarred Colli, who led A/B testing for Rocket Lawyer, puts it: "Shifting the discussion from 'What's testable?' to 'Everything is testable.'"

Adopting the mantra "Always Be Testing" is one of the tenets of taking your testing program long-term.

Redesigning Your Redesign: CareerBuilder and the Optimizely Website

One of the mistakes we have seen companies make is undertaking a complete redesign of their site and *then* optimizing the new site using A/B testing. This is a violation of two core principles of A/B testing: *define success metrics* and *explore before you refine*.

The failure to define success metrics comes as a result of redesigning the site *without a goal*. Other than wanting to look more current, what are the specific targetable behaviors or actions you want the redesigned site to encourage in your users?

The failure to explore before they refine comes when companies pass up their biggest opportunity to get hugely meaningful data from their redesign: testing the new design itself. Instead of comparing the new page against a tweaked version of itself, test it against the old site. The bigger the change you're making, the more you want to be sure that it's having a positive effect. *Then* you can worry about refining from there, once you've rolled out the new design with *evidence* that it's doing a better job than the old one in the critical areas.

David Harris, Senior Internal Business Systems Analyst at CareerBuilder, talks about how the long-term adoption of A/B testing at CareerBuilder has involved what he calls "pulling testing ever farther up in the process": moving it from deployment to design. "It is during the very early stages of redesigning a page that we are incorporating in plans to test things—how we want to test them, what we want to test for, what conversions are important."

During the summer of 2012, we at Optimizely were planning a total website redesign. We felt there was an opportunity for us to hone our look and our message for new visitors. What better

way to improve our own website than to eat our own dog food? We ran Optimizely on Optimizely.

Our previous site design made a clear and simple statement about our initial offering, an easy-to-use product for website A/B testing which we captured with our tagline: "A/B Testing You'll Actually Use." We discussed the opportunity we had to re-imagine a design that would maintain our core message and brand while providing additional benefits to visitors to our homepage and more broadly capture the value of A/B testing.

One other area of focus with the new site was providing a more comprehensive approach to engaging different visitor types. We have continued to learn over time how different customer types use Optimizely to achieve a wide range of different goals. To address this, we built out a series of pages that focused on the benefits of using Optimizely for each of these groups: agencies, developers, e-commerce sites, large "enterprise" sites, publishers, and small businesses.

Applying the approach of refine, explore, refine, our design team took to developing several concepts that could possibly achieve a new look and feel that could take our site experience to the next level in several areas. Ultimately, we agreed that the most important thing a potential tester can do on our site is enter a URL and experience our WYSIWYG editor firsthand. We focused our design efforts and our testing success metrics around maximizing the number of users who used our editor, and who signed up for an account. In addition to the key macro-conversion goals, it's important to track a wide range of goals to get a holistic sense of how the new site design performs.

In just about every metric we measured, the new site was a clear winner against the old one. After running the test on new visitors for just over a month, we were confidently able to declare a winner and push the new site live to all visitors. Most of our

thoughts about what our original site lacked were proven true (Figures 10.1 through 10.8).

With a new design live, we can dive into a range of new tests already lined up as we refine from here. There is still plenty of great work to be done.

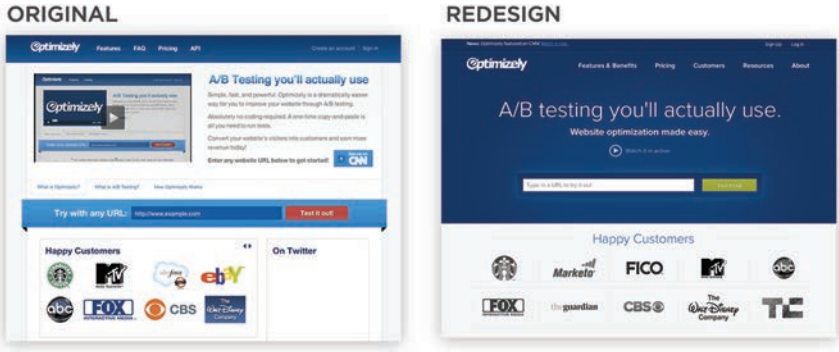

FIGURE 10.1 Original Optimizely homepage versus redesigned homepage.

FIGURE 10.2 Original Optimizely About page versus redesigned About page.

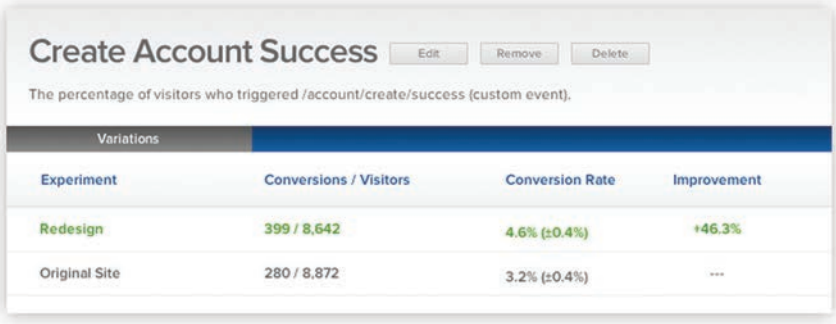

FIGURE 10.3 Create Account Success: Percentage of visitors who completed the "create an account" signup form.

FIGURE 10.4 Uses Editor: percentage of visitors who entered a URL on the homepage and used the Optimizely editor.

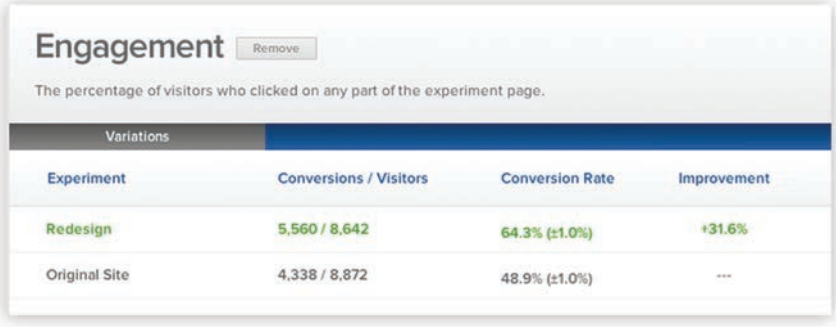

FIGURE 10.5 Engagement: this goal is included in every experiment created with Optimizely by default. It measures the percentage of visitors who click anywhere on the experiment page.

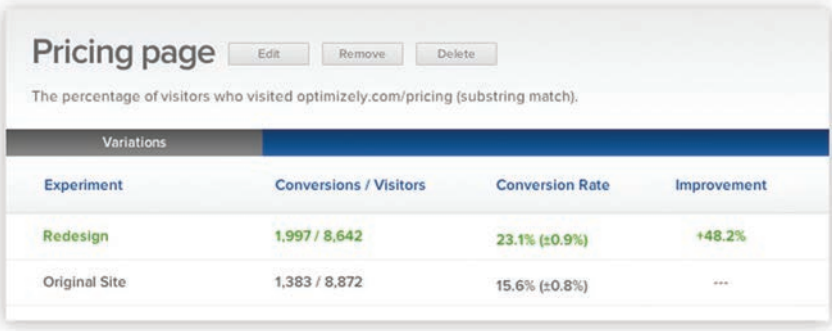

FIGURE 10.6 Visits Pricing: percentage of visitors who visited the Optimizely Pricing page.

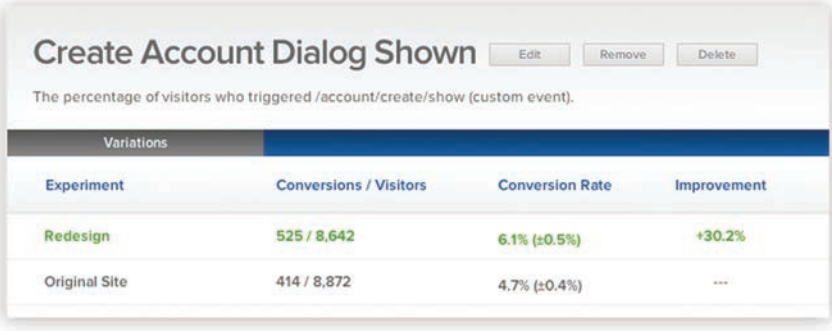

FIGURE 10.7 Create Account Dialog Shown: percentage of visitors who submitted the free trial signup form on the homepage.

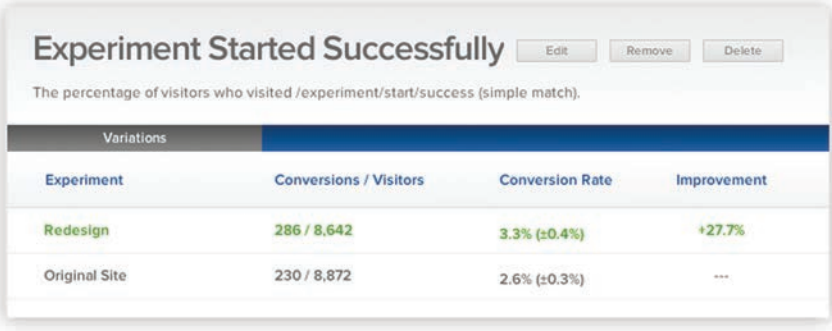

FIGURE 10.8 Experiment Started Successfully: percentage of visitors who landed on the experiment creation success confirmation page.

TL;DR

- **Multivariate tests** are a powerful way to test a number of variables simultaneously, and can reveal **interaction effects** between them. However, they require more traffic and can be slower to achieve statistical significance.
- We have found that the companies that have had the greatest success with A/B testing favor a nimbler, more **iterative approach** that tests a handful of different variants of a single variable at a time and incorporates the winner as they go on to the next test.
- When working on a complex change like a site redesign, we recommend that you **move testing further up the process** so that it becomes something that happens *during* the design and rollout of the new site, rather than *after*.