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

CHAPTER

13

Personalize, Personalize, Personalize

Moving Beyond the One-to-Many Web

Online retail giant Amazon has an army of engineers whose job it is to deliver an experience that maximizes profit per square inch for each page on the Amazon website. And they've realized that one of the best ways to do this is through personalization. They examine a user's entire history—the things someone has looked at, clicked on, or purchased—and then deliver exactly the thing they think is most valuable for *this* user.

Today's A/B testing tools offer advanced targeting and segmentation, allowing the non-Amazons of the world to create tailored experiences for different types of visitors. A business may start by using A/B testing to improve the "average best experience" on its website—a single experience across all types of users. The next step in the company's evolution of optimization is moving from honing the "average best experience" for everyone to grouping users into segments—and then using A/B testing to optimize the experience for each segment. A returning visitor could see a different interface than a new visitor; someone on a tablet could see something different from someone on a desktop; someone from Canada could see something different from someone from the United States. Each segment is an opportunity for optimization.

Targeting versus Segmentation

There are two things that Optimizely and other testing solutions enable to help sites make sense of their user populations: targeting

and segmentation. *Targeting* happens *before the test* and is essentially the definition of who is allowed to see a particular experiment, based on the URL and any number of conditions. One cohort of visitors (say those who come to your site via social media) will see one variation while visitors coming from search engines will see a different one.

Segmentation happens *after the test* and takes a different approach: you run the experiment for everybody and then isolate different groups of segments *afterward* and figure out how each performed. A great example of where segmentation can be handy is with mobile and tablet browsers. Perhaps your proposed new page layout works great overall: you still want to make sure that it's not a *regression* or an inferior experience for users arriving on smartphones or tablets. For instance, perhaps a button element is too small to be comfortably clicked on a mobile screen: this would likely show up in a segmentation report, where the conversion and engagement of mobile users is lagging behind that of their desktop counterparts.

A growing number of websites are moving from providing one-to-many, average best experiences for all visitors to *one-to-few* experiences that involve smart targeting based on browser, location, behavior, and more. The final step in this evolution is the *one-to-one* web: a personal experience tailored for each individual user. It's easy to imagine that several years from now people will look back and be shocked at how generic, impersonal, and one-to-many the web we know today is. It's inevitable that the web will move toward a more personalized one-to-one experience and it's just a question of how that reality will come about. The vision for Optimizely is to enable businesses to show exactly the right thing to the right person at the right time. A/B testing is the first step toward this vision. Targeting and segmentation are the next step.

Using Segmentation to Drill Down into Test Results

Imagine that a company tries a new site layout and notices that overall sales increase. A slightly more detailed analysis might reveal that the sales decreased slightly for people who visited the site on tablet devices. The company then starts hypothesizing about why that happened and what was different for this visitor segment. The issue could be, for instance, that the new design pushed the "Add to Cart" button below the fold on an iPad and the user had to scroll down to tap it. That's a good hypothesis about what could be causing sales to drop, and a perfect place to begin a second experiment. The company targets a segment or audience (perhaps just iPad users) and tries to deliver an even better experience to them.

The natural testing cycle is usually to start off by examining the one-to-many results data: how did all website visitors react to the test? Then, slicing the data into granular pieces based on any number of conditions (UTM source, browser, cookies, referral URL, location) generates specific questions like, how did paid traffic react to the test? Results from specific segments in one experiment turn into the targeting criteria for the next experiment. Iterative loop closed!

Geo-Targeting, State by State: Romney 2012

From the start, the digital campaign team for Mitt Romney's 2012 presidential campaign considered increasing email signups on mittromney.com to be one of their primary goals. As Ryan Meerstein—a senior political analyst from Targeted Victory who ran testing and optimization for the Romney campaign—explains, "Email is still the golden goose of fundraising when

you're making direct solicitations. We're seeing each email valued at anywhere between seven and eight dollars in future revenue."

Between May 2011 and November 2012, the Romney campaign's 140-person digital team along with Targeted Victory ran hundreds of tests. "Once we saw [how easy it was to conduct A/B testing], the ideas started flying. We wanted to start testing just about everything," Meerstein says. "We started on the splash page and when we saw success, we continued to build from there."

They tried showing different landing pages to visitors from different states, hypothesizing that visitors would sign up for email updates more if they saw a message specific to their state. They tested the state-specific geo-targeted message against a universal landing page lacking any state-specific messaging (Figure 13.1).

They found that when they simply added the state name (in the example in the figure, Florida) to the call to action text, visitors entered their email and zip code *19 percent more often*.

Starting in September 2012 and lasting until Election Day, visitors to mittromney.com received a distinct experience depending on their home state, which proved to be a valuable tool for the campaign. The data clearly showed that personalizing the message led to success. With this test as testament, the team decided to make the splash page specific for each state. They used geo-targeting to send visitors from each state to a page with a message specific to that state. They also crafted *personalized calls to action* based on absentee-vote states and early-vote states. Visitors from Ohio saw messages directing them to early voting locations; visitors from Colorado saw targeted messages for how to get an absentee ballot.

As a result, the Romney team saw not only greater signups on the splash page but more interaction with *local events* advertised on

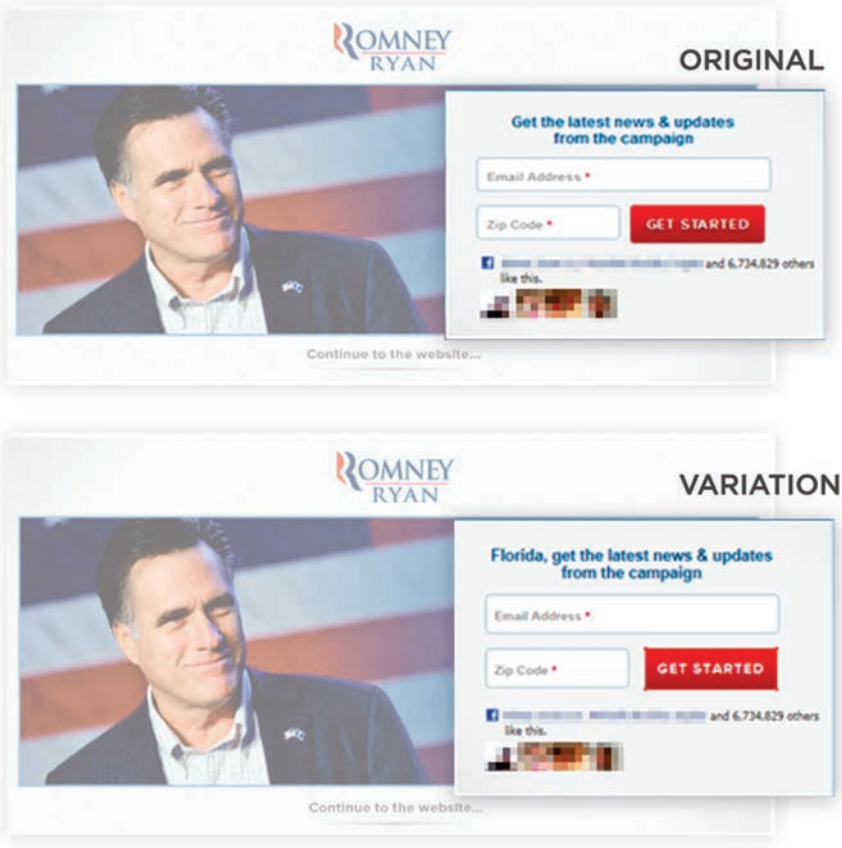

FIGURE 13.1 Nationwide Romney 2012 email signup page versus state-specific Romney 2012 email signup page.

Source: Romney 2012.

the site, *especially* in the critical hours after voting started. "The thing that was great about it was that we could go [to our A/B testing tool] and set up the personalized experiences in thirty minutes," Meerstein says. "In the final weeks of the campaign, there's a huge difference between something being live on Tuesday morning and Thursday night."

When to Personalize and When *Not* to: Wikipedia

Wikipedia is a useful case study in the art of *not* tailoring the experience to every user. For instance, when Wikipedia's annual fundraising push comes around, the site is able to know that one user may be coming from Toronto for information about film directors, and another from Sydney for information about world history, but it shows both users (and *all* users) the very same message. That's a deliberate choice, and it also happens to be the *correct* one. How do they know this? They've tested it.

Through A/B testing, the fundraising team discovered that their users appear to be *less* likely to donate when faced with a targeted fundraising appeal than with a universal one. It could be that users find personalization intrusive, or it could be that part of Wikipedia's brand is its openness and universality.

The Wikipedia example illustrates that despite its incredible power as a tool, personalization isn't *always* going to be more effective. Sometimes the "average best" is in fact simply "best." The web is not a place for static universal truths, and personalization is no exception. How is a company to know whether personalization will take its success metrics to the next level, or set them back? There is only one answer: *test it*.

TL;DR

- **Segmentation** allows you to compare how different segments of users responded to the same experience. Differences that you observe between groups can be illuminating and give you an opportunity to go beyond the **average best experience** of the one-to-many web toward an improved one-to-few experience.
- **Targeting** is deliberately providing different types of users with different experiences, and can be a powerful technique under many circumstances.
- Consider how your user experience may or may not be optimized across different **platforms**, screen dimensions, touch input versus mouse input, and so on.
- Sometimes **geo-targeting**, or tailoring users' experience based on their location, can be an extremely powerful way to optimize beyond the "average best."
- While personalization is frequently a boon for your success metrics, occasionally a **universal** call to action works better than a targeted one. Test to make sure personalization is working for your key success metrics.