

Coconut: Behavioral Economics for Website Optimization

Master of Information Management & Systems Final Project 2012

UC Berkeley School of Information

<http://groups.ischool.berkeley.edu/coconut/>

Evan M. Smith - evan@ischool.berkeley.edu

Arthur A. C. Suermondt - arthur@ischool.berkeley.edu

Jeff L. Zych - jlzych@ischool.berkeley.edu

Advisor: Steven Weber - steve@ischool.berkeley.edu

Introduction

In recent years A/B testing has become a powerful and popular tool to test the efficacy of user interfaces. Although software has been developed to overcome the technical challenges of administering tests (such as setting them up and analyzing the results), little has been done to aid users in forming hypotheses, i.e. why is one interface better than another and therefore worth testing. Although a strength of A/B testing is that tests are cheap and easy to deploy, and therefore anything can be tested, this is not always ideal. Unless a site has significant traffic, tests can take weeks to get statistically significant results, making it both unfeasible and unwise to test every design variation. Testing designs that are more likely to achieve a site's goals will yield greater returns more quickly.

To overcome this problem, we propose that insights from the field of behavioral economics can provide web developers a stepping stone when designing A/B tests and forming hypotheses. Broadly speaking behavioral economics is the study of how people make decisions. Using the web can be viewed as a series of decision making steps, ranging from low-level (should I click this link?) to high-level decisions (should I buy this product?). With this view in mind, we used behavioral economic phenomena to develop concrete suggestions of tests users can run on their site. However, given the vastness of the web, and the numerous goals that exist across sites, we decided to focus our suggestions on non-profits and increasing donation amounts. We then embodied these suggestions in a web-based tool that walks users through using these suggestions on their site.

Usability Testing Shortcomings

In recent years A/B testing has become a powerful and popular tool to test the efficacy of user interfaces. By running experiments to determine which design choices (e.g. blue or red buttons, bold or italic fonts) influence website visitors' behavior, organizations are able to quickly improve their websites. The tools that are used to support such testing focus on automating the process of administering the tests but do not provide much in the way of supporting advanced hypothesis formation. Therefore, the A/B tests run are purely inductive and do not provide a disciplined means by which to maximize the effect of the testing.

Meanwhile, there have been many findings in the behavioral economics and psychology literature that if applied to the design of websites could have profound impacts and influence visitor behavior in significant ways. To our knowledge, there are no A/B testing tools that integrate these findings directly

into their platforms. Our belief is if the experimental evidence that has come forth from academic studies were made more readily available to web designers, in turn the designers could perform more advanced hypothesis formation that would ultimately result in more meaningful improvements to their websites in a faster and more efficient means.

Project Scope

Our project aims to assist web designers in modifying their website such that they are able to influence the behavior of their visitors. We have selected a handful of major behavioral economics findings to include in our tool. Our system embodies these findings into usable portions of website functionality that can be inserted into one's website with minimal effort. Although our tool does not administer the A/B tests, our system integrates with A/B testing platforms such that the web designers can immediately take the new website functionality that we suggest and test its effect on influencing user behavior.

For our tool's initial release we have focused on serving non-profit organizations who use their website as a means to solicit monetary support for their mission. Oftentimes, non-profit organizations have less time to spend on website optimization related matters due to limited resources (money, time, personnel). In turn, our tool can help fill this void by augmenting the limited efforts that do exist. Our focus on the non-profit domain does not limit the applicability of our tool to a broader audience. For example, monetary donations collected on non-profit websites often use the same mechanics as traditional shopping carts used more broadly on the web. In turn the use cases that we handle in great detail can be abstracted to apply to a broader target audience in the future.

System Overview

Our system is structured as an assistant to the web designer that takes as input inventory of the organization's overarching strategic goals and outputs suggested A/B tests. At the onset of a session, the user is prompted to enter the address of the website they are going to improve, then the assistant asks multiple-choice questions of the user to better understand the organizational context and intended goals of their website. Our system maps the collected data with appropriate suggestions from our library of suggested improvements. The suggestions were developed by us based on findings from behavioral economics, and are described in more detail below. In addition, our assistant provides brief explanations of the underlying theories that are informing these suggestions.

Non-Profit Topology

To support our tool’s suggestion engine, we have developed a multifaceted organizing system to classify non-profits. Our team performed an inductive exercise whereby we reviewed the top 25 non-profit websites by total private donation funding in 2010 (Forbes, 2011). This exercise resulted in four facets by which we classify non-profits (see table 1). We believe the variation along the following dimensions has a major influence on the structure and content of the organization’s web presence and therefore has a significant effect on the structure and content of our suggestions.

Facet	Probing Question	Example
How Impact is Made	Does the non-profit deliver programs directly to beneficiaries or does the non-profit perform general research?	Direct = Habitat for Humanity that builds homes for people that need them General Research = Cancer Society that performs general research
Goals for Website	What are the conversion goal(s) of the organization’s website?	Monetary donations, signing up volunteers, convincing people to join the organization
How Impact is Communicated on the Site	Does the non-profit use messaging that demonstrates that supporting donations assist targeted or named beneficiaries or does the messaging not specify any targeted beneficiary?	Targeted = Kiva.org where you specify to whom you wish to direct your donation Non-Targeted = Salvation Army where the organization collects donations for a general fund
Tone and Imagery of Messaging	Does the organization demonstrate the struggle that they aim to address or rather focus on the positive impact of the organization?	GrassrootSoccer.org is focused on a positive image, showing the impact they have WorldVision.org emphasizes the struggle they aim to address

Table. 1 - *The typology we have created to classify non-profit websites that use our assistant.*

Suggestion Engine

We have created a library of suggestions that have been informed by the behavioral economics literature that we have reviewed. These suggestions include a short description of what to change, an example (with code snippets, when necessary), and a brief (1 to 3 paragraph) explanation of the

underlying theories (with links to more information). At run time, using the data gathered from the assistant, we classify organizations and their websites using the above facets and provide suggestions that trace to these organizational characteristics. For example, our ‘unit-based donation’ suggestion requires the charity to provide good or services directly to those in need and have the goal of collecting donations from site visitors. In this case, the assistant gathers what issue the organization addresses and then uses that to provide the appropriate art and copy for the issues that the non-profit addresses. If the non-profit addresses health issues the website copy will include health related donation items. Other suggestions developed for the prototype are described in Table 2, but many more can thought of based on the behavioral economics theories discussed in the next section.

Suggestion	Classification
<i>Unit Based Donation</i> - donation amounts are specified using a an interactive widget that calculates total donation amount by the user’s inputted unit amount selection	How Impact Is Made = Directly to Beneficiaries How Impact is Communicated = Targeted Goal = Donations
<i>Priming Survey</i> - before asking the visitor for a donation, the site presents the user with a two choice survey with the following question “Which of the following two charitable goods would have the biggest impact”	How Impact Is Made = Directly to Beneficiaries, Goal = Donations
<i>Defaults</i> - suggests an optimal default donation amount based on the current setting or lack of setting	Goal = Donations
<i>Channel Factor</i> - provides template of clear instructions that enable potential volunteer to visualize how they would easily participate	Goal = Volunteer

Table. 2 - Our suggestion engine maps user inputs to relevant suggestions from our library . The suggestions listed above are not exhaustive and just represent a sample.

Applying Behavioral Economics to the Problem Space

Broadly speaking, behavioral economics is the study of how people make decisions. In contrast to classical economics, which assumes people are rational and have perfect information during the decision making process, behavioral economics takes findings from the field of psychology and cognitive science to explain the times people make decisions that are “irrational,” such as when emotions or imperfect heuristics are used (Tversky & Kahneman, 1974).

Viewing the web as a series of decision making processes, the behavioral economic findings can apply to web design. The decisions on the web range in complexity and granularity from low-level (should I click this link? Which form option should I choose?) to high-level decisions (should I buy this product? Should I sign up for this newsletter?).

Following are several behavioral economic phenomena and explanations of how they can be applied to the web design domain. We have used these principles to inspire the content in our suggestion library.

Defaults

When people are given a choice the majority are likely to stick with the default option. This has been shown to hold across a wide variety of domains and situations. For example, most countries have an option for people to donate their organs upon their death. In America, the default choice is to not donate organs, i.e. people must specifically check a box on a form (also called an “opt-in” system), and the consent rate is only about 28%. In contrast, Belgium’s default option is to donate organs (also called an “opt-out” system), in which about 98% of the population consent to donation. (Johnson and Goldstein, 2003)

Rationally speaking, there shouldn't be any difference in what people choose based on the default — they should choose whatever benefits them most. But as this example shows, that is not the case. One reason for this is that making a choice takes effort (even if it’s only checking a box), whereas sticking with the default is cognitively effortless. Additionally, when decisions have unclear costs and benefits that are difficult to evaluate (such as organ donation), defaults can imply the organization’s “recommended” option, thus saving people the time of thinking through the choice themselves.

Porting this to information systems is fairly straightforward: applications and services should have well-thought-out default preferences, and forms should have default options marked. As software studies have shown, most users never change the default preferences. For example, less than 5% of people who use Microsoft Word ever change any settings (Spool, 2011). The implications for designers are twofold: first, good defaults provide a better user experience by saving users the effort of configuring software (or marking choices on forms); and second, they provide a way to direct users towards selecting options

that are most in line with the designer's (or organization's) goals. However, the strength of this effect is weaker the easier the choice is, e.g. the default will have almost no effect when the user is choosing the country they live in from a drop down menu (since this information is well-known to them and takes almost no cognitive effort to figure out), as opposed to Facebook's privacy settings (which are not as clear and the effects of each option are difficult to reason out a priori). Although this has become more well known recently in the design world, it is not yet ubiquitous.

Anchoring

Anchoring is another well-studied, robust phenomenon summarized as, "different starting points yield different estimates, which are biased toward the initial values" (Tversky and Kahneman, 1974). In other words, if a person already has a number in their head (the "anchor"), values they produce thereafter will be unconsciously adjusted from this starting point. For example, one study asked people about their willingness "to save 50,000 offshore Pacific Coast seabirds from small offshore oil spills". Both groups were asked how much they would be willing to contribute, but one group was first asked if they'd be willing to pay \$5. The group that was anchored to \$5 had average contributions of \$20, whereas the group with no anchor averaged at \$143 (Kahneman, 2011).

This can be applied online to anything that asks users to input a number, such as donations to charity or item quantities in web stores. Many charities have variations of "as little as \$10 dollars can make a difference!" on their donate page, which undoubtedly influences donation amounts. Experimenting with this value could potentially have a meaningful impact on the donation amounts received.

Framing

Framing is the simple idea that different ways of presenting the same information will evoke different emotional reactions, and thus influence people's decision making processes. A simple example is the statement "the odds of survival one month after surgery are 90%" is more reassuring than the equivalent statement of "mortality within one month of surgery is 10%." One study that explored how this phenomenon affects behavior showed two groups of women videos on breast cancer and mammography in order to convince them to get screened. The first group's video was gain-framed, espousing the benefits, whereas the second group's video was loss-framed and emphasized the risks. Only 51.5% of those who saw the gain-framed video got a mammography, compared to 61.2% of those who saw the loss-framed video (Banks et al., 1995). The information presented was equivalent, but the

way the information was framed greatly influenced the outcome.

Marketers and advertisers have used this technique for a long time to influence people's purchasing decisions, both online and offline. For example, many products claim to "Kill 99.9% of germs" instead of its converse, "Only 0.1% of germs survive!". But this technique can be utilized by any organization that's trying to spur people to action (such as getting mammograms or volunteering for a cause), and they should take care to optimally frame their message.

Priming

The classic psychology definition of priming describes the phenomenon whereby exposing a person to a stimulus (image, word, idea, etc.) will influence their response to a later stimulus, in an automatic, subconscious way (Kolb & Whishaw, 2003). As applied to behavioral economics, making a thought salient in someone's mind will affect their decision making process. For example, one study showed that a proposition to increase school funding received significantly more votes when the polling station was at a school than when it was at other locations. The key idea here is the imagery of a school triggered school-related norms, such as supporting children's futures, which subconsciously influenced people to support the proposition (Berger, Meredith and Wheeler, 2008).

Applying this to the digital realm, websites should use imagery positively associated with their cause to subtly influence visitors. Once again this is pretty standard in marketing and advertising, but can also be applied to charities, politics (both for controversial issues and politicians), and advocacy sites. For example, Barack Obama's site (barackobama.com) contains images of him giving speeches, smiling, shaking hands, etc., on almost every page, keeping the image of him as a strong, confident leader in people's minds. This subtly reinforces these associations in people's minds to increase donations and volunteer sign-ups (Segall, 2012).

Choice Sets

Another well-studied topic is choice sets, the basic idea is that too many options increases the cognitive costs of comparing and evaluating options, resulting in suboptimal decision making (or the abandonment of choosing altogether). The classic study that defined this concept presented jam samples at supermarkets with the intention of enticing people to buy jam. When 24 were available to try, only 3% of people who stopped bought a jar of jam. But when just 6 were available, 30% of people

made a purchase (Botti & Iyengar, 2006).

Consumers browsing web stores face the same difficulties as those browsing jam in the above example: too many product offerings with unclear differentiators leads to choice paralysis. Companies should offer clear, simple choices whenever possible. However, this also applies to more than just sales and marketing. For example, organizations that offer multiple volunteering opportunities should make sure to have a few distinct possibilities.

At an even lower level, too many buttons, navigation elements, and links on a page leads to confusion among users. Similarly, long forms with lots of options (i.e. in drop down menus, radio buttons, and checkboxes) have low conversion rates (Wroblewski, 2008). Web designers should ensure to have a limited number of distinct options in navigation menus and forms.

Empirical Testing of Our Suggestions

To begin the process of validating the efficacy of our suggested designs, we ran an A/B test on our 'unit-based donation' suggestion. We asked participants for basic demographic information and their historical donation habits (i.e. frequency and amount). The participant was then told they had a limited amount of funds left over from taxes with which they could donate to charity. We then told them they would have the option of donating to 5 charities and presented them with one of two pages: a standard donate form, which had a list of radio buttons and donation amounts; or our 'unit-based donation' form (see fig. 1), which asks users to donate a number of 2-by-4s at \$10 a piece (or they may enter a different, arbitrary amount). In actuality only one charity is presented, and the test ends after the donation transaction is completed.

Coconut Communities

Coconut Communities' mission is to build affordable housing for families in need. We receive donations of building materials as donations to keep our cost as low as possible. At the core of our mission is our belief that stable and comfortable homes are essential to the functioning of a family unit. Wooden 2-by-4 beams cost \$10 each and are essential for constructing a home. A typical house needs about 200 2-by-4s to complete.

Make a donation

Choose from one of these common donation amounts:

- \$ 10
 \$ 20
 \$ 50
 \$ 100
 \$ 200
 Or, enter other amount:

Your donation will be:

\$ 0

Coconut Communities

Coconut Communities' mission is to build affordable housing for families in need. We receive donations of building materials as donations to keep our cost as low as possible. At the core of our mission is our belief that stable and comfortable homes are essential to the functioning of a family unit. Wooden 2-by-4 beams cost \$10 each and are essential for constructing a home. A typical house needs about 200 2-by-4s to complete.

Donate 2-by-4 beams

Select the number of 2-by-4 beams you would like to donate:

0 wooden 2-by-4 beams x \$10 each = \$ 0

Donate Additional Funds

Amount (\$):

Your donation will be:

\$ 0

Fig. 1 - The two conditions of the experiment. Radio-button condition on the left, unit-based condition on the right.

Both conditions were presented with the following introductory text upon seeing the first (and only) charity, 'Coconut Communities':

"Coconut Communities' mission is to build affordable housing for families in need. We receive donations of building materials as donations to keep our cost as low as possible. At the core of our mission is our belief that stable and comfortable homes are essential to the functioning of a family unit. Wooden 2-by-4 beams cost \$10 each and are essential for constructing a home. A typical house needs about 200 2-by-4s to complete."

During the creation of our experiment the following considerations guided our thinking. Due to the fact that users are not actually donating money we needed to instill a sense of scarcity in participants to simulate real conditions. In order to do this we told users they had a limited amount of funds after taxes and purposefully did not specify an amount. If we had specified an amount during this part of the experiment, participants would have anchored on that number, thus minimizing the differences between the two interfaces. We further primed participants before this step by asking them to explicitly think about their donation habits, and to enter how much they typically donate. Next, during the experiment we told users they would be presented with 5 charities so that they wouldn't donate all of their money to just the first and only charity that we presented to them. In other words, we wanted users to think critically about whether they really wanted to donate to the one charity we showed them, rather than automatically donating to complete our survey.

Experiment Design and Results

We recruited 115 participants by inviting them to fill out our survey via e-mail, Facebook and personal requests. On average, participants donated more in the unit-based condition ($M = 55.00, SE = 5.86$), than in the radio-button condition ($M = 42.17, SE = 9.80$). This difference, however, was not significant $t(76) = -1.191, p > .05$ (two-tailed $p = .237$). Since this experiment was merely conducted as a plausibility probe we were not expecting the results to be significant, but only an indication of whether our suggestions could be successfully applied in a real world environment.

Before analyzing the results using any statistical methods we removed participants who donated zero dollars (15 in the radio-button condition, 21 in the unit-based condition) or who indicated that on average they donate over 5,000 dollars (three outliers of \$5,000, \$10,000 and \$35,000) from the data set. Since our conversion metric is a donation amount, implying that a person actually donates, participants who donated zero dollars were not relevant for the comparison of mean donation amounts (this is equivalent to someone getting to the donate form and then leaving the page without donating on an actual site). They do, however, provide an interesting indication that there were possible other shortcomings in our experiment design, although it could also be that these participants simply did not relate to the charity that we presented.

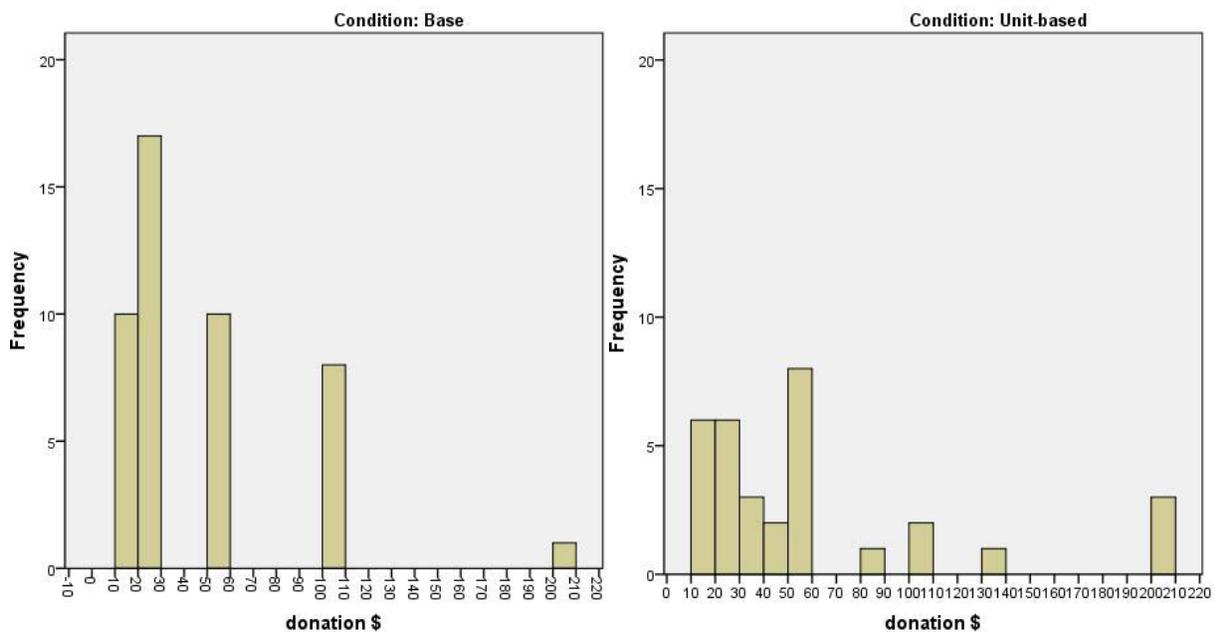


Fig. 3 - Frequency distribution of donation dollar amounts for the radio-button condition (left) and the unit-based condition (right).

Another finding worth mentioning is the observed difference in frequency distributions of dollar amounts between the radio-button condition and the unit-based condition. The unit-based condition, which featured a slider and calculated dollar amounts, appears to have a smoother distribution in the popular range of \$10 - \$50 dollars (78.13% of all donations), with donations for every \$10 step in between. The radio-button condition, on the other hand, only has donation for \$10, \$20 and \$50 in this range (80.44% of all donations). Additionally, the unit-based condition has donations in four outliers (\$80, \$100, \$130 and \$200, 21.88% of all donations) whereas the radio-button condition only has donations in two outliers (\$100, \$200, 19.56% of all donations). We recognize that the radio-button condition was set up in a way that would likely yield these results. At the same time, it confirms one of the assumptions behind the unit-based suggestion, which is that users want to be able to adjust their donation amount in a more finely controlled way.

Lessons Learned from Our Experiment

Though our library of suggestions has been directly inspired by behavioral economics phenomena, the designs we have created are not guaranteed to render a significant, or even positive result. Even though the scale of this experiment was limited and the results not statistically significant, we are pleased with the result of a 30.4% increase in donation dollars in the unit-based condition versus the radio-button condition (mean difference = -12.83). This outcome warrants the further exploration of website design patterns inspired by the theories from behavioral economics.

Conclusions and Potential Revenue Models

This project presented an attempt to bridge the gap that exists between running an A/B test, and knowing what changes to actually test. To do this, we took findings from the field of behavioral economics and applied them to the websites of charities. Our tool works by classifying non-profits according to a typology we developed via an analysis of charity sites, which is then used as a basis to provide suggestions. We developed the suggestions based on findings in the behavioral economics literature, and included examples (with code snippets) to aid implementation alongside short descriptions of the underlying theories to explain why we think a change will work. Although we were unable to test all of our suggestions, the one we did test provided initial validation of our idea (but was not statistically significant). After showing suggestions, it is up to the user to actually implement them and run the tests using an external service.

We don't provide an A/B testing platform ourselves. Instead, our assistant refers users to our partners, such as Optimizely and Visual Website Optimizer (web-based visual A/B testing suites). This also provides opportunities for revenue models. Optimizely, for example, offered to provide our customers a discount when using their tool. This discount could easily be used as a commission for our referral, thereby generating revenue whenever users start a new experiment from a Coconut suggestion. Other possibilities include sponsored suggestions, where third parties would provide suggestions in addition to Coconut's own suggestions. In the non-profit space this could include suggestions from other online platforms such as Change.org. Overall, we think that our strategy of demonstrating the impact of our tool will convince users to come back and try out new suggestions, establishing a sustained source of revenue.

Future Directions

Going forward there are numerous directions our work could go. To start, we've barely scratched the surface of possible suggestions. Besides the myriad of variations one can make on our initial set of suggestions, there are many more theories from behavioral economics that could form the basis of improvements. Next, it would be beneficial to actually test our suggestions in more real-world scenarios. Although they are based on well-studied theories, that does not guarantee they will be impactful. Additionally, we would have liked to do user testing to demonstrate the efficacy and usefulness of our tool, and make improvements where necessary. Finally, we chose to focus our project on one specific market: non-profits. Expanding into other markets is an obvious direction to go, but requires careful study and understanding of that vertical. Even given these shortcomings, we are confident our work validated our initial concept of applying findings from behavioral economics to website optimization.

References

- Banks, S. M., Salovey, P., Greener, S., Rothman, A. J., Moyer, A., Beauvais, J., & Epel, E. (1995). The effects of message framing on mammography utilization. *Health Psychology, 14*(2), 178-184. doi:10.1037/0278-6133.14.2.178
- Berger, J., Meredith, M., & Wheeler, S. C. (2008). Contextual priming: Where people vote affects how they vote. *Proceedings of the National Academy of Sciences of the United States, 105*(26), 8846-8849. doi:10.1073/pnas.0711988105
- Botti, S., & Iyengar, S. S. (2006). The dark side of choice: When choice impairs social welfare. *Journal of Public Policy and Marketing, 25*(1), 24-38. doi:10.1509/jppm.25.1.24
- Forbes. (2011). *The 200 largest U.S. charities*. Retrieved April, 27, 2012, from http://www.forbes.com/lists/2011/14/200-largest-us-charities-11_rank.html
- Johnson, E. J., Goldstein, D. (2003). Do defaults save lives? *Science, 302*(5649), 1338-1339. doi:10.1126/science.1091721
- Kahneman, D. (2011). Anchors. In *Thinking, fast and slow* (Kindle ed., pp. 124-125). New York, NY: Macmillan - Farrar, Straus and Giroux.
- Kolb, B., & Whishaw, I. Q. (2003). Chapter 16: The frontal lobes. In *Fundamentals of human neuropsychology* (pp 453-454, 457). New York, NY: Worth Publishers.
- Segall, L. (2012, April 24). Optimizely aims to give Obama 2012 a data edge. *CNN Money*, <http://money.cnn.com//2012/04/23/technology/startups/optimizely-election/index.htm>
- Spool, J. (2011, September 14). Do users change their settings. Message posted to <http://www.uie.com>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. doi:10.1126/science.185.4157.1124
- Wroblewski, L. (2008). *Web form design: Filling in the blanks*. New York, NY: Rosenfeld Media.