

RAPID DESIRABILITY TESTING:

A Case Study

Michael Hawley, Senior Vice President, Director of Experience at Mad*Pow

THERE CAN OFTEN BE DISAGREEMENTS AMONG THE MEMBERS OF A PROJECT TEAM ON WHICH DESIGN DIRECTION WE SHOULD CHOOSE. AS PRACTITIONERS OF USER-CENTERED DESIGN, IT IS NATURAL FOR US TO TURN TO USER RESEARCH TO HELP INFORM AND GUIDE THE PROCESS OF CHOOSING A VISUAL DESIGN.

In the design process we follow at my company, Mad*Pow Media Solutions, once we have defined the conceptual direction and content strategy for a given design and refined our design approach through user research and iterative usability testing, we start applying visual design. Generally, we take a key screen whose structure and functionality we have finalized—for example, a layout for a home page or a dashboard page—and explore three alternatives for visual style. These three alternative visual designs, or comps, include the same content, but reflect different choices for color palette and imagery.

The idea is to present business owners and stakeholders with different visual design options from which they can choose. Sometimes there is a clear favorite among stakeholders or an option that makes the most sense from a brand perspective. However, there can often be disagreements among the members of a project team on which design direction we should choose. If we've done our job right, there are rationales for our various design decisions in the different comps, but even so, there may be disagreement about which rationale is most appropriate for the situation.

As practitioners of user-centered design, it is natural for us to turn to user research to help inform and guide the process of choosing a visual design. But traditional usability testing and related methods don't seem particularly well suited for assessing visual design for two reasons:

1. When we reach out to users for feedback on visual design options, stakeholders are generally looking for large sample sizes—larger than are typical for a qualitative usability study.

2. The response we are looking for from users is more emotional—that is, less about users' ability to accomplish tasks and more about their affective response to a given design.

With this in mind, I was very intrigued by recent posts about desirability testing from Christian Rohrer on his [xdStrategy.com](#) blog. In one entry, Christian posits desirability testing as a mix of quantitative and qualitative methods that allow you to assess users' attitudes toward aesthetics and visual appeal. Inspired by his overview of this method, we researched desirability studies a bit further and tried a modified version of the method on one of our projects. This article reviews the variants of desirability testing that we considered and the lessons we learned from conducting a desirability study to assess the visual design options for one of our projects.

WHY IS DESIRABILITY IMPORTANT?

From a usability perspective, an important role of visual design is to lead users through the hierarchy of a design as we intend. Use of value contrast and color and the size and placement of elements can serve to support a product's underlying information architecture and interaction design. During the early stages of the design process, we focus on these functional aspects of a design and conduct research to ensure that the overall solution offers a compelling value proposition to users. We also aim to optimize usability and make it easy for users to realize the solution's benefits and, ultimately, achieve their goals.

A product's having valuable features and an intuitive information architecture and interaction design certainly contributes to

THE HALO EFFECT OF THAT EMOTIONAL RESPONSE CAUSES USERS' FIRST IMPRESSIONS OF A DESIGN TO IMPACT A PRODUCT'S OR APPLICATION'S PERCEIVED UTILITY, USABILITY, AND CREDIBILITY.

its overall desirability. However, there is a difference between functional desirability and the emotional desirability that stems from aesthetics, look, and feel. Visual elements can support a solution's interaction design, but they can also elicit an emotional response from users. Understanding and exploiting these emotional responses can help designers to influence users appropriately.

Interestingly, Lindegaard and his associates found that a design can have an emotional impact very quickly. In their research report "Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression!" they outline a series of experiments they conducted to assess how quickly people form an opinion about the visual appeal of a design. As you can probably guess from the title of their report, they found that a design elicits an emotional response very rapidly—in about the time it takes to read a single word.

This is important because the halo effect of that emotional response causes users' first impressions of a design to impact a product's or application's perceived utility, usability, and credibility. Users generally form their first impressions less by interacting with certain functions and more through their initial emotional response to a product's visual aesthetics and imagery. Researchers classify the effects as positive or negative. For example, if a user has a positive first impression of the design aesthetics, they are more likely to overlook or forgive poor usability or limited functionality. With a negative first impression, users are more likely to find fault with an interaction, even if a product's overall usability is good and the product offers real value.

This has special implications for a number of domains. For example, in an ecommerce environment, a site's perceived level of trustworthiness can affect buying decisions or people's willingness to interact with the site. For interactive applications, a sense of

organization can affect perceived usability and, ultimately, users' overall satisfaction with the product.

SO WHY NOT JUST ASK PEOPLE WHICH DESIGN THEY LIKE BETTER?

As I noted earlier, within my company's design process, we try to iteratively improve our conceptual approaches and interaction designs through user feedback and usability testing. Often, during this testing, we use a think-aloud protocol and ask participants to explain which option they prefer for an interaction and why. With visual design comps, it is tempting to simply show participants the design options at the end of a usability test session and ask them which they like better. This sounds straightforward enough and, generally, we've found that this is what business stakeholders think of when we talk about getting user feedback on visual designs.

The problem with this simplistic approach is that people's rationales for the overwhelming variety of their tastes may or may not be related to the business or brand goals for a design. For example, when I've asked this question before, I've heard participants say they like a certain design because it's "their favorite color" or "I like things that are green." Their statements may be truthful, but those types of responses don't help researchers assess the emotional impact of a design or how it aligns with the intended brand attributes. In addition, some participants have a difficult time articulating what it is about a design they like or dislike. During an interview, participants may be able to select a preferred design, but without a structured mechanism for providing feedback, they may be at a loss for words when it comes to describing why they like or dislike it.

We've also found that, when asking for design preferences during a qualitative study like a usability test, the small sample sizes do not align with stakeholder expectations for

validation of a given design. Especially for public-facing Web sites and applications, their visual design is one of the most significant depictions of the company's brand, and business sponsors and stakeholders often want substantial customer feedback to assure them a given direction is correct.

SOME POTENTIAL RESEARCH METHODS

Besides simply asking for users' preferences for particular designs, we explored several other structured research methods that could help inform design selection, including the following:

- triading
- experience questionnaires
- quick-exposure memory tests
- measurement of physiological indicators

Triading

The triading method I described in one of my columns on UXmatters offers potential in this regard, because it is structured around the comparison of several options. The idea with triading is to elicit attributes that research participants and target users would use to compare given alternatives, in a way that is not biased by the researcher. Given three design options, a researcher could ask participants to identify two that are different from the third and describe why they are different. This process helps the researcher to understand what dimensions are important to target users in comparing different designs. We've found this method to be very helpful both when evaluating the competitive landscape and for assessing different conceptual options from an interaction design perspective. However, this method is difficult when conducting studies with large sample sizes, and it can be difficult to present the tabulation of results to stakeholders who are looking for research to help them choose the best design option.

Experience Questionnaires

Another possible approach to assessing design options is a comprehensive experience questionnaire. Questionnaires such as SUS (System Usability Scale), QUIS (Questionnaire for User Interface Satisfaction), and WAMMI (Website Analysis and Measurement Inventory) are broad, experience-based questionnaires, but do include questions relating to visual appeal and aesthetics. In a 2004 report to the Usability Professionals' Association, "A Comparison of Questionnaires for Assessing Website Usability," Tom Tullis and Jacqueline Stetson wrote about a study that compared the effectiveness of these questionnaires. They found that, to varying degrees, all of these questionnaires were effective in reliably assessing differences between Web sites.

For comparing visual design options, questionnaires' ability to identify perceived differences between design alternatives is intriguing. These questionnaires are also attractive, because they are relatively straightforward and easy to administer on a large scale. But many of the questionnaires also include a significant number of questions about interactivity and require participants to have had a certain level of interaction with a site or application. For a quick comparison of static visual design comps, we felt these questions would not be appropriate. In addition, we were not just looking for a winner among the designs, we wanted to understand what emotional responses each alternative elicited, so we could make better design decisions going forward. The output of these questionnaires did not lend itself to that purpose.

Quick-Exposure Memory Tests

A third approach we looked at was a quick-exposure memory test. In this method, researchers show participants a user interface for a very brief moment, then take it away. Then, they ask participants to recall what

THE TRIADING METHOD ...
IS STRUCTURED AROUND THE
COMPARISON OF SEVERAL
OPTIONS.
FOR COMPARING VISUAL
DESIGN OPTIONS,
QUESTIONNAIRES' ABILITY
TO IDENTIFY PERCEIVED
DIFFERENCES BETWEEN
DESIGN ALTERNATIVES IS
INTRIGUING.

RESEARCHERS SHOW PARTICIPANTS A USER INTERFACE FOR A VERY BRIEF MOMENT, THEN TAKE IT AWAY. THEN, THEY ASK PARTICIPANTS TO RECALL WHAT THEY REMEMBER ABOUT THE USER INTERFACE FROM THAT BRIEF EXPOSURE.

“IN RESEARCHING POTENTIAL METHODS FOR DESIRABILITY TESTING, WE REVIEWED THE GROWING BODY OF KNOWLEDGE ABOUT THE PHYSIOLOGICAL INDICATORS RESEARCHERS CAN MEASURE TO ASSESS EMOTIONAL RESPONSE.”

they remember about the user interface from that brief exposure. Participants have limited interaction with the site or application, so theoretically, they’re providing you a glimpse into their first impression—what sticks in their memory. During usability test sessions, we’ve tried this method to elicit conversation about home pages and other starting pages, and it is helpful in assessing layout considerations and information design.

There is a service available online called [fivesecondtest](#) that lets you solicit responses from visitors and get a decent sample size—that is, 50 participants—in a relatively short period of time. We chose not to use this service as our primary method for visual design comparison studies, because we felt it focused too much on people’s memory of particular items rather than emotional impact, but for a small amount of money and effort, it may be helpful in certain situations.

Measurement of Physiological Indicators

Finally, in researching potential methods for desirability testing, we reviewed the growing body of knowledge about the physiological indicators researchers can measure to assess emotional response. In the article “A Multi-method Approach to the Assessment of Web Page Designs,” Westerman and his co-authors summarize the available approaches:

- **Electroencephalography (EEG)** measures activity in parts of the brain that you can map to certain emotional responses.
- **Electromyography (EMG)** measures muscle activity that correlates to excitement levels.
- **Electrodermal Activity (EDA)** measures the activity of sweat glands, which is said to correlate to arousal and excitement.
- **Blood Volume Pressure (BVP)** measures dilation in the blood vessels, which, in turn, correlates with arousal.
- **Pupil dilation** appears to correlate to both arousal and mental workload.

- **Respiration** measurements can indicate negative valence or arousal.

Similar to eyetracking, during these studies, various sensors track these physiological measurements as researchers show participants particular designs. Changes in one or more indicators suggest a particular emotional response. Researchers often pair these measurements with attitudinal and self-reporting surveys to give a multifaceted view of participants’ emotional reactions to a design. The potential of these physiological methods of quantitatively assessing emotional response is great. However, because of the time and budget constraints on many of our projects, we were looking for an approach we could use outside a lab or even over the Internet, so we could get large samples of responses.

OUR PREFERRED METHOD FOR ASSESSING THE DESIRABILITY OF VISUAL DESIGNS

Of all the methods we’ve considered, the one that seemed to align best with our goals was the approach Joey Benedek and Trish Miner of Microsoft described in their paper “Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting.” Working collaboratively with a multidisciplinary team, Benedek and Miner developed a set of adjectives research participants could use to describe their reactions to a user interface. They put all of these adjectives, shown in Figure 1, on product reaction cards with which participants could interact. But the important part is that they developed a list of terms that were potential descriptors of the user interface and were also potentially salient for their research. These adjectives represented a mix of descriptions that people might consider positive or negative. They showed participants a user interface, then asked them to select the three to five of these adjectives they thought best described it.

By analyzing the resulting data across participants, researchers can align certain adjectives with each visual design option and assess how each option aligns with a business's intended emotional response and brand attributes. Researchers can use this method in either a one-on-one setting or a survey. The advantage of the one-on-one approach is that the researcher can probe participants' rationales for why they chose certain

The complete set of 118 Product Reaction Cards		
Accessible	Creative	Fast
Advanced	Customizable	Flexibl
Annoying	Cutting edge	Fragile
Appealing	Dated	Fresh
Approachable	Desirable	Friendl
Attractive	Difficult	Frustr
Boring	Disconnected	Fun
Business-like	Disruptive	Gets ir
Busy	Distracting	Hard to
Calm	Dull	Helpfu
Clean	Easy to use	High q
Clear	Effective	Imper
Collaborative	Efficient	Impre
Comfortable	Effortless	Incom
Compatible	Empowering	Incons
Compelling	Energetic	Ineffe
Complex	Engaging	Innov
Comprehensive	Entertaining	Inspiri
Confident	Enthusiastic	Integr
Confusing	Essential	Intimic
Connected	Exceptional	Intuiti
Consistent	Exciting	Invitin
Controllable	Expected	Irrelev

Fig. 2 Microsoft product reaction cards

adjectives and potentially uncover additional insights. Obviously, with a survey-based study, researchers would miss the qualitative aspects of a one-on-one study, but they would gain the impact of a larger sample size. Either way, the structured aspect of the study makes data analysis relatively straightforward. Additionally, reporting participants' top adjectives for each design option to various stakeholders is both impactful and easy to comprehend.

OUR EXPERIENCE

We tried this approach to desirability testing on a recent project to see whether it would help us refine our visual design direction for a public-facing Web site. Once we'd reached the point in our overall design process where we'd finalized the content, messaging, and information hierarchy, we started designing multiple visual concepts for the site.

The goal of the site was to persuade customers to sign up for a discount health plan that could offer them savings on out-of-pocket medical expenses. Our goals for the site's design and emotional impact were as follows:

- We wanted to portray a professional and trustworthy image to overcome any objections consumers might have if they weren't familiar with the brand.
- We didn't want a site that would appear gimmicky or overly promotional and discourage customers.
- We sought to design a site that potential customers would find friendly and genuinely approachable.
- Given the sensitive nature of healthcare expenditures, we wanted visitors to feel comfortable with the site and let a sense of empathy come through the design.

With these goals in mind, we developed two alternative visual design options. In the first option, shown in Figure 2, we used clean edges and bold colors in an effort to make the site appear conservative and stable. Our assumption was that visitors might find similarities between this site and other well-known brands with which they are familiar. This, in turn, would help them develop a sense of trust in the site. In the second design, shown in Figure 2, we opted for a softer, warmer color palette, with rounded corners and welcoming images to give the site a friendly feel.

BY ANALYZING THE RESULTING DATA ACROSS PARTICIPANTS, RESEARCHERS CAN ALIGN CERTAIN ADJECTIVES WITH EACH VISUAL DESIGN OPTION AND ASSESS HOW EACH OPTION ALIGNS WITH A BUSINESS'S INTENDED EMOTIONAL RESPONSE AND BRAND ATTRIBUTES.

WE TRIED THIS APPROACH TO DESIRABILITY TESTING ON A RECENT PROJECT TO SEE WHETHER IT WOULD HELP US REFINE OUR VISUAL DESIGN DIRECTION FOR A PUBLIC-FACING WEB SITE.

TO TEST WHICH APPROACH WOULD BEST ALIGN WITH OUR INTENDED GOALS, WE CONDUCTED A DESIRABILITY TEST USING PRODUCT REACTION CARDS.

WE IDENTIFIED THE ADJECTIVES PARTICIPANTS SELECTED WITH THE HIGHEST FREQUENCY AND TALLIED THE TOTAL NUMBERS OF POSITIVE AND NEGATIVE ADJECTIVES FOR EACH DESIGN.

To test which approach would best align with our intended goals, we conducted a desirability test using product reaction cards. Starting with the full Microsoft list of cards, we revised the list to include only the adjectives we felt were important for this brand, after assessing our early user research. We narrowed the final list to 60 adjectives, but kept the 60/40 split between positive and negative terms Benedek and Miner had suggested.

We conducted the study through a survey, dividing participants into three groups. We showed the first group only the first design

our client was keen on our asking the simple preference question, so we decided to do so. Finally, we gave all participants an opportunity to comment on and give their rationale for their adjective choices or preferences. Through our survey, we collected responses from 50 people in each of the three group.

As we expected, the results from the third group were inconclusive. Participants in that group were evenly divided in their preferences and their rationales for their decisions varied widely. However, tabulating the adjectives the other two groups had selected from the list proved

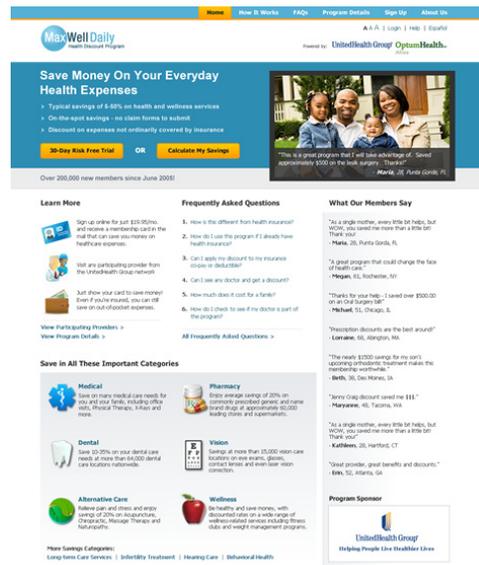


Fig. 2 Visual Design Option 1

option, instructing them to select five adjectives from the list that they thought best described the design. We showed the second group only the second design option, giving them the same instructions. Because the designs were static screenshots, participants were not able to interact with either of them. We showed the third group both design options—alternating which design we showed participants first to minimize order bias—and asked which design they preferred. We had hypothesized that data analysis of the results from the third group would be difficult, but



Fig. 3 Visual Design Option 2

to be very helpful. We identified the adjectives participants selected with the highest frequency and tallied the total numbers of positive and negative adjectives for each design.

Contrary to our assumptions before conducting this research, while participants thought the first option was both understandable and clear, they also described it as sterile, sophisticated, and impersonal. The sense of trustworthiness we had intended did not come across as one of the adjectives for that design. As we had anticipated, participants saw the second option as approachable and friendly, but surprisingly,

they also described it as professional and trustworthy. Obviously, all of these adjectives were in line with our intended emotional response. Additionally, the second option received a much higher percentage of positive adjectives than the first option.

Compared to the simple Which design do you like better? question, our survey of product adjectives did a much better job of informing and helping us to achieve consensus on our design decisions. Based on our research findings and a review of participant comments, we developed consensus between designers



Fig. 4 Our Final Design

and business stakeholders, selecting the second design option as the starting point for design refinements. Best of all, when others outside the project team questioned the appropriateness of a design element, because they liked other styles, we were able to provide a research-based rationale that minimized preference disagreements and moved us toward successful completion of the project.

CONCLUSION

The prospect of trying to measure people's emotional responses to different visual design options, then choose the best design can often be daunting. Everyone has a different opinion, and wading through volumes of data on simple preferences seems counterproductive. Plus, research that measures people's emotional responses to a design is complex in nature. Their experiences of a visual design are multifaceted, and a number of different design aspects can impact their response to a product. Measurement of physiological responses to designs shows promise as a means of assessing people's overall emotional reactions to a product, but not everyone has access to labs and measurement devices.

The design-adjective approach to desirability studies I've reviewed here is both easy to implement and helpful in isolating the emotional impact of a visual design. My company has now used this method several times, and we've been pleased with the clarity the results have provided. Not only have our desirability studies helped us to select a design direction, the insights we've gained from our research have challenged our assumptions as designers and informed our revisions of our chosen design direction.

Add desirability testing to your research toolkit. Then, the next time a senior executive on a project says, "Make it purple—that's my daughter's favorite color!" desirability testing can save the day!

RESOURCES

Benedek, Joey, and Trish Miner.

"Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting." Proceedings of UPA 2002 Conference, Orlando, FL, July 8-12, 2002. Retrieved February 10, 2010.

Lindgaard, Gitte, Gary Fernandes, Cathy Dudek, and J. Brown.

"Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression!" Behaviour and Information Technology, 2006. Retrieved February 10, 2010.

Rohrer, Christian.

"Desirability Studies: Measuring Aesthetic Response to Visual Designs." xdStrategy.com, October 28, 2008. Retrieved February 10, 2010.

Tullis, Thomas, and Jacqueline Stetson.

"A Comparison of Questionnaires for Assessing Website Usability." Usability Professionals' Association Conference, 2004. Retrieved February 10, 2010.

Westerman, S. J., E. Sutherland, L. Robinson, H. Powell, and G. Tuck.

"A Multi-method Approach to the Assessment of Web Page Designs." Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, 2007. Retrieved February 10, 2010.

ABOUT MAD*POW

Mad*Pow is an experience design agency that specializes in designing and developing intuitive user experiences for internal systems, customer facing applications, marketing sites, e-commerce sites and interactive media.

We are located in the heart of the New Hampshire Seacoast in historic Portsmouth, NH.

Since 2000, we have been providing award-winning creative ideas and personalized account service for industry leaders and large organizations such as Google, Monster, Aetna, Fidelity, Timberland, Intuit, The Hartford and McKesson.

Mad*Pow is comprised of 29 talented and passionate user experience professionals, each with a deep focus on helping our clients achieve their business goals and produce measurable results.

