Email Optimization: How A/B testing generated $500 million in donations
Introduction
You’ve heard this story...

Analysis: Obama won with a better ground game
By Rebecca Sinderbrand, CNN
updated 7:00 AM EST, Wed November 7, 2012
Fact check: TRUE

- The Ground Game
  - 786 Obama field offices vs. 284 Romney offices
  - In Ohio:

- Thousands of paid staffers organized 2 million volunteers to get voters registered and to the polls.

Source: TheMonkeyCage.Org
You’ve heard this story…

How Obama's data crunchers helped him win

Built to win: Deep inside Obama’s campaign tech

Friended: How the Obama Campaign Connected with Young Voters

How the Obama Campaign's Data-Miners Knew What You Watched On TV

Hacking the Electorate: Data's role in Changing Minds
Fact check: TRUE

Sophisticated technology and targeting
- Giant data & micro-targeting operation
- Fully integrated databases
- Smarter new methods of targeting TV ads
- New technology developed in-house for social sharing, polling place lookup, phone banking, volunteer mobilization, vote tracking, election day rapid response
What you didn’t hear

All of that costs money
The fundraising challenge

- In 2008, Obama campaign raised $750 million
- Would not be enough in 2012

$750 million?
Not impressed.
The fundraising challenge

- But fundraising was proving more difficult in 2012 than in 2008
  - President less available for fundraising events
  - In early campaign, we saw average online donation was half of what it had been in 2008
  - People were giving less, and less often

- We had to be smarter, and more tenacious
The fundraising challenge

- The real game-changer? It’s something most of you already do everyday.

A / B

- More A/B testing than any campaign ever
First, the results:

- Raised more than half a billion dollars online
- 4.5 million donors
- $53 average gift
How did we do it?

Lessons Learned:

1. Content matters (…and so do subject lines).
2. Don’t trust your gut.
3. Being pretty isn’t everything.
4. Incentives work.
5. Invest in your team.
6. Foster a culture of testing.
What impact can testing have?

<table>
<thead>
<tr>
<th>THE SUBJECT LINE</th>
<th>THE HAUL</th>
<th>THE DIFFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The team tested numerous subject lines</td>
<td>Based on the donations</td>
<td>They then projected how much less money the campaign would collect if they used anything other than the most successful e-mail</td>
</tr>
<tr>
<td>by sending fundraising e-mails to small</td>
<td>those e-mails raised, the</td>
<td></td>
</tr>
<tr>
<td>groups of supporters</td>
<td>team projected how much</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the pitches would bring in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>if sent to the full Obama</td>
<td></td>
</tr>
<tr>
<td></td>
<td>list</td>
<td></td>
</tr>
<tr>
<td>I will be outspent</td>
<td><strong>$2,540,866</strong></td>
<td>n/a</td>
</tr>
<tr>
<td>Some scary numbers</td>
<td>$1,941,379</td>
<td>$599,487</td>
</tr>
<tr>
<td>If you believe in what we’re doing…</td>
<td>$911,806</td>
<td>$1,629,060</td>
</tr>
<tr>
<td>Last call: Join Michelle and me</td>
<td>$894,644</td>
<td>$1,646,222</td>
</tr>
<tr>
<td>Would love to meet you</td>
<td>$755,425</td>
<td>$1,785,441</td>
</tr>
<tr>
<td>Do this for Michelle</td>
<td>$714,147</td>
<td>$1,826,719</td>
</tr>
<tr>
<td>Change</td>
<td>$711,543</td>
<td>$1,829,323</td>
</tr>
<tr>
<td>The most popular Obama</td>
<td>$659,554</td>
<td>$1,881,312</td>
</tr>
<tr>
<td>Michelle time</td>
<td>$604,813</td>
<td>$1,936,053</td>
</tr>
<tr>
<td>Deadline: Join Michelle and me</td>
<td>$604,517</td>
<td>$1,936,349</td>
</tr>
<tr>
<td>Thankful every day</td>
<td>$545,486</td>
<td>$1,995,380</td>
</tr>
<tr>
<td>The one thing the polls got right…</td>
<td>$403,603</td>
<td>$2,137,263</td>
</tr>
</tbody>
</table>

**Subject:** I will be outspent

The winning subject line wound up beating the projections: it raised **$2,673,278**.
Testing = constant improvement

- Little improvements add up
  - Improving 1% here and 2% there isn’t a lot at first, but over time it adds up
Test every element

- Question: what footer language should we use to reduce unsubscribes?

<table>
<thead>
<tr>
<th>Variation</th>
<th>Recips</th>
<th>Unsubs</th>
<th>Unsubs per recipient</th>
<th>Significant differences in unsubs per recipient</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Email 1" /></td>
<td>578,994</td>
<td>105</td>
<td>0.018%</td>
<td>None</td>
</tr>
<tr>
<td><img src="image2" alt="Email 2" /></td>
<td>578,814</td>
<td>79</td>
<td>0.014%</td>
<td>Smaller than D4</td>
</tr>
<tr>
<td><img src="image3" alt="Email 3" /></td>
<td>578,620</td>
<td>86</td>
<td>0.015%</td>
<td>Smaller than D4</td>
</tr>
<tr>
<td><img src="image4" alt="Email 4" /></td>
<td>580,507</td>
<td>115</td>
<td>0.020%</td>
<td>Larger than D3 and D4</td>
</tr>
</tbody>
</table>

*Note: The email templates are examples and should be adapted to fit the specific campaign requirements.*
Tests upon tests upon tests

- Every piece of communication is an opportunity to test
- A single email can have many tests attached

- Subject & draft tests
- Full-list tests
- Background personalization tests
No, really. Test every element.

- Running tests in the background via personalized content

---

From: Barack Obama <info@barackobama.com>
Subject: My last campaign

David --

In a few days, I'll be hitting the trail for my last campaign.

Everything we've accomplished in the past three years -- and our chance to do so much more -- is on the line.

What we do today will be a measure of whether or not we're ready to fight for it.

**Donate $61 or whatever you can before tonight's fundraising deadline.**

By pitching in before midnight, you'll automatically be in the running to join me and George Clooney at his place on May 10th.

It's not often I can get away from work, so I look forward to spending a fun evening in L.A. with a couple supporters like you.

In the meantime, let's close out this deadline strong:
Longitudinal tests

- Example: how much email should we send?

- Experiment: gave sample audience higher volume of email for an extended time, and compared to control

- Results: More email = more donations
  - People may say they get too much email
  - But mild annoyance proved to be the worst result
  - Unsubscribes accrued linearly
  - Donations did, too.

- Implementing a “more email” policy probably led to $20-30 million in additional revenue for the campaign.
So we made shirts.
Lesson #1

Drafts and Subject Lines Matter
Example: Draft language

Amelia --

If every supporter of this campaign steps up when it counts, we will win this election.

From now until November 6th, every single day is going to count.

We have an important fundraising deadline at midnight, and we saw this week how much is at stake in this election.

Will you step up?

https://donate.barackobama.com/Midnight-Deadline

Let's win.

Barack
Example: Subject lines

Test sends

<table>
<thead>
<tr>
<th>version</th>
<th>Subject line</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1s1</td>
<td>Hey</td>
</tr>
<tr>
<td>v1s2</td>
<td>Two things:</td>
</tr>
<tr>
<td>v1s3</td>
<td>Your turn</td>
</tr>
<tr>
<td>v2s1</td>
<td>Hey</td>
</tr>
<tr>
<td>v2s2</td>
<td>My opponent</td>
</tr>
<tr>
<td>v2s3</td>
<td>You decide</td>
</tr>
<tr>
<td>v3s1</td>
<td>Hey</td>
</tr>
<tr>
<td>v3s2</td>
<td>Last night</td>
</tr>
<tr>
<td>v3s3</td>
<td>Stand with me today</td>
</tr>
<tr>
<td>v4s1</td>
<td>Hey</td>
</tr>
<tr>
<td>v4s2</td>
<td>This is my last campaign</td>
</tr>
<tr>
<td>v4s3</td>
<td>[NAME]</td>
</tr>
<tr>
<td>v5s1</td>
<td>Hey</td>
</tr>
<tr>
<td>v5s2</td>
<td>There won't be many more of these deadlines</td>
</tr>
<tr>
<td>v5s3</td>
<td>What you saw this week</td>
</tr>
<tr>
<td>v6s1</td>
<td>Hey</td>
</tr>
<tr>
<td>v6s2</td>
<td>Let's win.</td>
</tr>
<tr>
<td>v6s3</td>
<td>Midnight deadline</td>
</tr>
</tbody>
</table>

- Each draft was tested with three subject lines
- One subject line would usually be common across all drafts, to help make comparisons across messages
Example: Best vs. Worst Versions

Test sends

<table>
<thead>
<tr>
<th>version</th>
<th>Subject line</th>
<th>donors</th>
<th>money</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1s1</td>
<td>Hey</td>
<td>263</td>
<td>$17,646</td>
</tr>
<tr>
<td>v1s2</td>
<td>Two things:</td>
<td>268</td>
<td>$18,830</td>
</tr>
<tr>
<td>v1s3</td>
<td>Your turn</td>
<td>276</td>
<td>$22,380</td>
</tr>
<tr>
<td>v2s1</td>
<td>Hey</td>
<td>300</td>
<td>$17,644</td>
</tr>
<tr>
<td>v2s2</td>
<td>My opponent</td>
<td>246</td>
<td>$13,795</td>
</tr>
<tr>
<td>v2s3</td>
<td>You decide</td>
<td>222</td>
<td>$27,185</td>
</tr>
<tr>
<td>v3s1</td>
<td>Hey</td>
<td>370</td>
<td>$29,976</td>
</tr>
<tr>
<td>v3s2</td>
<td>Last night</td>
<td>307</td>
<td>$16,945</td>
</tr>
<tr>
<td>v3s3</td>
<td>Stand with me today</td>
<td>381</td>
<td>$25,881</td>
</tr>
<tr>
<td>v4s1</td>
<td>Hey</td>
<td>444</td>
<td>$25,643</td>
</tr>
<tr>
<td>v4s2</td>
<td>This is my last campaign</td>
<td>369</td>
<td>$24,759</td>
</tr>
<tr>
<td>v4s3</td>
<td>[NAME]</td>
<td>514</td>
<td>$34,308</td>
</tr>
<tr>
<td>v5s1</td>
<td>Hey</td>
<td>353</td>
<td>$22,190</td>
</tr>
<tr>
<td>v5s2</td>
<td>There won't be many more of these deadlines</td>
<td>273</td>
<td>$22,405</td>
</tr>
<tr>
<td>v5s3</td>
<td>What you saw this week</td>
<td>263</td>
<td>$21,014</td>
</tr>
<tr>
<td>v6s1</td>
<td>Hey</td>
<td>363</td>
<td>$25,689</td>
</tr>
<tr>
<td>v6s2</td>
<td>Let's win.</td>
<td>237</td>
<td>$17,154</td>
</tr>
<tr>
<td>v6s3</td>
<td>Midnight deadline</td>
<td>352</td>
<td>$23,244</td>
</tr>
</tbody>
</table>

Full send (in millions)

- $2.2 million additional revenue from sending best draft vs. worst, or $1.5 million additional from sending best vs. average
Some of the best subject lines:

- Hey
- Say you're with me
- Hell yeah, I like Obamacare
- Listen:
Lesson #2

Don’t Trust Your Gut
Testing = data-driven decisions

- We don’t have all the answers
  - Conventional wisdom is often wrong
  - Long-held best practices are often wrong
  - Going with things that had previously tested well was often wrong

- There was this thing called the Email Derby…
Lesson #3

The Prettiest Isn’t Always the Best
Experiments: Ugly vs. Pretty

- We tested sleek and pretty
  - That failed, so we asked: what about ugly?

- Ugly yellow highlighting got us better results
  - But at some point it lost its novelty and stopped working – always important to re-test!
Lesson #4

Incentives Matter
People respond to incentives

- Offering a free bumper sticker for enrolling in our Quick Donate program increased conversions by 30%
  - Our Quick Donate program, in turn, raised donation rates by 50% or more

- Giving away bumper stickers and car magnets, then daisy-chaining to a donate page, yielded enough donations to pay for the freebies immediately
Lesson #5

Invest In Your Team
OFA Digital Department

- Grew from a small team in spring 2011 to a department of 200+ in 2012
  - Outbound (email, social, mobile, blog)
  - Ads
  - Front-End Development
  - Design
  - Video
  - Project management
  - Digital Analytics
More voices, more talents

- **Outbound Team**
  - 18 email writers
  - 4 social media writers & bloggers

- **Digital Analytics Team**
  - 15 analysts with overlapping skills
    - Database management (SQL, Python)
    - Data analysis (Stata, R, SPSS)
    - Web analytics (Google Analytics, Optimizely)
The human element and our voice

Honesty

Authenticity
Real people, real characters

Ann Marie Habershaw
*Tough love*

Rufus Gifford
*An emotional roller coaster*
Lesson #6

Foster a culture of testing
The culture of testing

☐ Check your ego at the door

☐ Use every opportunity to test something

☐ Compare against yourself, not against your competitors or “the industry”
  ☐ Are you doing better this month than last month?
  ☐ Are you doing better than you would have otherwise?
Keep a testing calendar

- On the Obama campaign we had short-term and long-term calendars for national emails.
- We added a “tests” column to plan out which tests would be attached to which emails.
- If we saw blank spaces, it would remind us to think of more tests to run!
  - Important to do frequent brainstorming sessions.
Circulate your test results internally

- We had an internal listserv entirely for the express purpose of circulating test results

- Helped get buy-in and increased familiarity with the testing process

- Prompted discussions and generated new ideas for tests
The Big Picture
Testing wins.

- This mentality was applied across the board:
  - Helped recruit 2 million volunteers
  - Helped build for thousands of phone banks, rallies, and events
  - Got information and “the message” into the hands of our best messengers

- Did we mention raising half a billion dollars?
  - Testing resulted in about $200 million in additional revenue
    …and that’s a conservative estimate
Big data ≠ big brother

- Testing allows you to listen to your user base
  - Let them tell you what they like
  - Optimization gives them a better experience

- Usually, the interactions that are the most human are the ones that win
Experiments: Personalization

- Adding “drop-in sentences” that reference people’s past behavior can increase conversion rates
- Example: asking recent donors for more money

...it's going to take a lot more of us to match them.

Will you donate $25 or more today?

...it's going to take a lot more of us to match them.

You stepped up recently to help out -- thank you. We all need to dig a little deeper if we're going to win, so I'm asking you to pitch in again. Will you donate $25 or more today?

- Added sentence significantly raised donation rate
  - Confirmed in several similar experiments
Final Conclusions

- Big groups of smart people working together can accomplish a lot, even in 18 months

- But you don’t have to have a staff of hundreds to have a good testing program
  - Train existing staffers, hire more when you can
  - Foster a culture of testing: every piece of communication is an opportunity to test something
  - Even a small list can be split in two – do what you can