

Web Experiments

Neil Sengupta and Calvin Xavier Gomez

Learning Goals 🤖

- **A/B Testing**
- **Ramping Up and Auto abort**
- **The Facebook Experiment**
- **Debate on ethics**

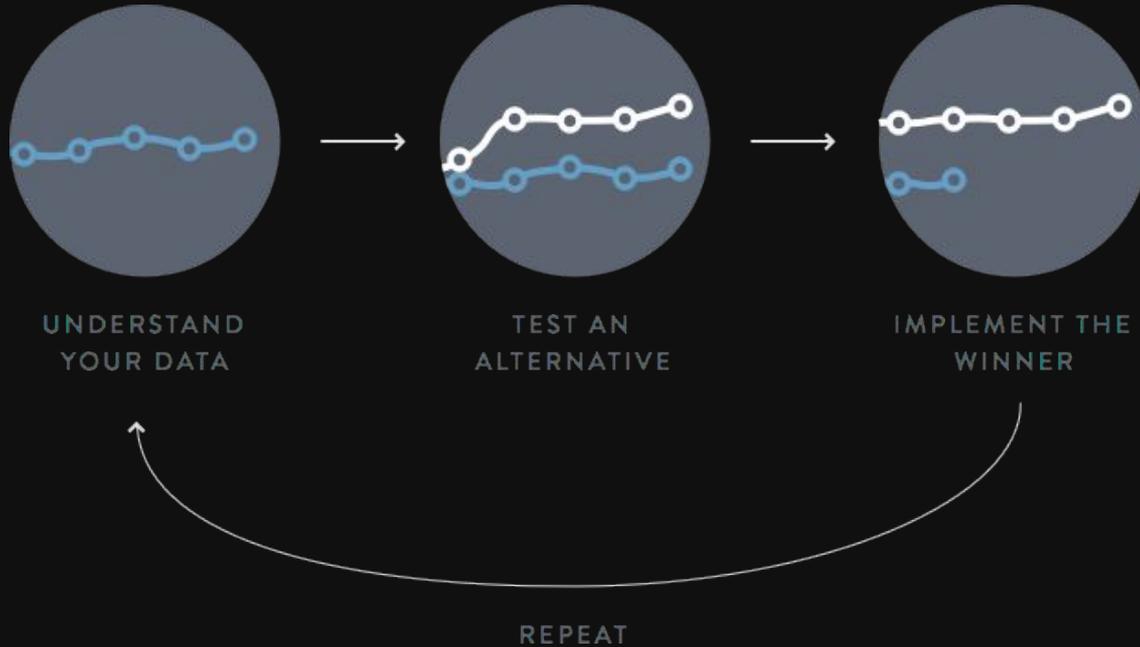
Controlled experiments on the web

Kohavi, Longbotham, Sommerfield, Henne

Key Ideas to take away from this paper

- A/B tests, Multivariate Tests, Control/Treatments tests, parallel flights.
- Practical Guide to conducting online experiments.
- Significant learning and increase in ROI when development teams listen to customers.

It's All A/Bout Testing



Thanks Obama!



The image shows a screenshot of the Obama '08 'GET INVOLVED' sign-up page. At the top center is the Obama '08 logo, consisting of a stylized 'O' with a blue and red wave pattern, and the text 'OBAMA'08' below it. Below the logo is a large blue banner with the text 'GET INVOLVED' in white, uppercase letters. The banner features a photograph of Barack Obama in a dark suit, looking to the right, surrounded by a crowd of people holding blue signs that say 'Obama '08'. Below the banner is a white section containing a sign-up form. On the left of this section is the text 'JOIN THE MOVEMENT'. To its right are two input fields: 'Email Address' and 'Zip Code'. To the right of these fields is a prominent red button with the text 'SIGN UP' in white, uppercase letters. At the bottom of the page, there is a dark blue footer with the text 'PAID FOR BY OBAMA FOR AMERICA' on the left, a small Obama logo in the center, and the text 'CONTINUE to WEBSITE' on the right.

Media

Button

SIGN UP NOW

LEARN MORE

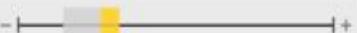
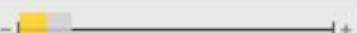
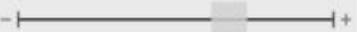
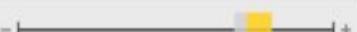
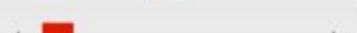
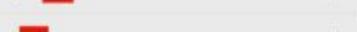
SIGN UP NOW

SIGN UP

Combinations (24)

Page Sections (2)

Download:  XML  CSV  TSV |  Print

| Relevance Rating  | Variation | Est. conv. rate  | Chance to Beat Orig.  | Observed Improvement  | Conv./Visitors  |
|--|-------------------|---|--|--|--|
| Button  | Original | 7.51% ± 0.2%  | — | — | 5851 / 77858 |
| | Learn More | 8.91% ± 0.2%  | 100% | 18.6% | 6927 / 77729 |
| | Join Us Now | 7.62% ± 0.2%  | 73.5% | 1.37% | 5915 / 77644 |
| | Sign Up Now | 7.34% ± 0.2%  | 13.7% | -2.38% | 5660 / 77151 |
| | Media | Original | 8.54% ± 0.2%  | — | — |
|  | Family Image | 9.66% ± 0.2%  | 100% | 13.1% | 4996 / 51696 |
| | Change Image | 8.87% ± 0.2%  | 92.2% | 3.85% | 4595 / 51790 |
| | Barack's Video | 7.76% ± 0.2%  | 0.04% | -9.14% | 3992 / 51427 |
| | Sam's Video | 6.29% ± 0.2%  | 0.00% | -26.4% | 3261 / 51864 |
| | Springfield Video | 5.95% ± 0.2%  | 0.00% | -30.3% | 3084 / 51811 |

Combinations (24)

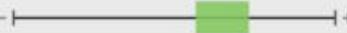
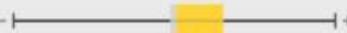
Page Sections (2)

Download:  XML  CSV  TSV |  Print

Disable

All Combinations (24) ▾

Key:  Winner  Inconclusive  Loser 

| <input type="checkbox"/> Combination | Status  | Est. conv. rate  | Chance to Beat Orig.  | Observed Improvement  | Conv./Visitors  |
|--|--|---|--|--|--|
| Original | Enabled | 8.26% ± 0.5%  | — | — | 1088 / 13167 |
|  Top high-confidence winners. Run a follow-up experiment » | | | | | |
| <input type="checkbox"/> Combination 11 | Enabled | 11.6% ± 0.6%  | 100% | 40.6% | 1504 / 12947 |
| <input type="checkbox"/> Combination 7 | Enabled | 10.3% ± 0.6%  | 100% | 24.0% | 1340 / 13073 |
| <input type="checkbox"/> Combination 3 | Enabled | 9.80% ± 0.6%  | 99.7% | 18.7% | 1277 / 13025 |
| <input type="checkbox"/> Combination 10 | Enabled | 9.23% ± 0.6%  | 95.9% | 11.7% | 1203 / 13031 |
| <input type="checkbox"/> Combination 8 | Enabled | 9.03% ± 0.6%  | 91.6% | 9.28% | 1178 / 13046 |
| <input type="checkbox"/> Combination 9 | Enabled | 8.77% ± 0.6%  | 81.8% | 6.10% | 1111 / 12672 |
| <input type="checkbox"/> Combination 6 | Enabled | 8.64% ± 0.5%  | 75.3% | 4.58% | 1108 / 12822 |



OBAMA'08

CHANGE

WE CAN BELIEVE IN



**JOIN THE
MOVEMENT**

Email Address

Zip Code

LEARN MORE

PAID FOR BY OBAMA FOR AMERICA



CONTINUE  WEBSITE

Mailchimp's A/B Test



AB Testing

Help

Save and Exit ▾

What would you like to test?

Select the variable you want to test. We'll generate a campaign for each combination of the variable, up to 3 combinations.

+ Subject Lines

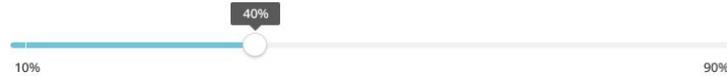
+ From Names

3
Content
- +

+ Send Times

How would you like to test the combinations?

- Distribute the combinations across all recipients
- Test on a percentage of recipients to send a winner



The remaining 60% of your recipients will be sent the winning combination.

What metric should determine the winner?

By Click Rate ▾ after hours ▾

We recommend waiting at least 4 hours to gather all the results.

Summary



| | |
|---|--------|
| Total recipients | 37,500 |
| Test segment | 40% |
| Combination size | 5,000 |
| We recommend at least 5,000 recipients per combination. | |
| Winning segment | 60% |
| Remaining recipients | 22,500 |

Mailchimp's A/B Test

The screenshot displays the Mailchimp A/B testing interface. At the top, the campaign is titled "Fall Shirts" and the current variant is "Content 2". Navigation options include "Help", "Preview and Test", "Save as Template", and "Save and Exit".

The main content area shows a preview of the email. The subject line is "Shirts on shirts on shirts!". The main image features a man with a beard wearing a blue denim shirt against a blue background. Below the image is the headline "Fall Shirts Are Here" and a paragraph of text: "'Tis the season at J.R. Flourence for fitted long sleeve button downs! Our full Autumn Series of artisanal, handmade poly cotton blends is debuting in stores this weekend, but mailing list members get early online access." A dropdown menu is open over the text, showing three options: "Content 1", "Content 2" (selected), and "Content 3".

On the right side, there is a "Content" panel with a grid of design elements: Text, Boxed Text, Divider, Image, Image Group, Image Card, Image + Caption, Social Share, Social Follow, Button, Footer, and Code.

At the bottom, a navigation bar shows "Content" as the active tab, with "Template", "Design", and "Description" as other options. "Back" and "Next" buttons are also present.

Mailchimp's A/B Test

Fall ShirtsHelpSave and Exit

Content setup

Test different messages or templates within your campaign.

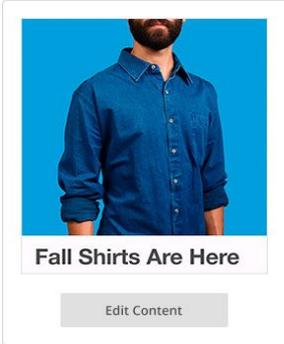


Fall Shirts Are Here

Edit Content

1 Column - Banded Template
Edited on Aug 07, 2015 09:48 am

"Testing the order in which the content appears: main alpha image (red), headline,"

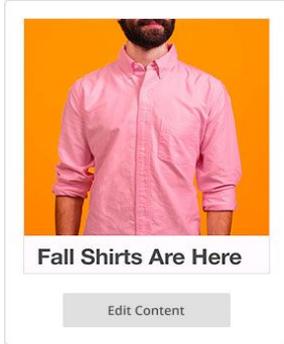


Fall Shirts Are Here

Edit Content

1 Column - Banded Template
Edited on Aug 07, 2015 09:37 am

"Testing the order in which the content appears: main alpha image (blue), headline,"



Fall Shirts Are Here

Edit Content

1 Column - Banded Template
Edited on Aug 07, 2015 09:53 am

"Testing the order in which the content appears: main alpha image (yellow), headline,"

Mailchimp's A/B Test

The winning combination is scheduled to send on Aug 07, 2015 05:46 pm
[Convert to manual selection](#)

Early Online Access List

Export

15,000 Recipients

List: Early Online Access

Tested on: 40% of 37,500 subscribers

Variates tested: Content

Subject: Fall Shirts Are Here

Subscribers per combination: 5,000

Delivered: August 6, 2015, 3:00pm

Winning metric: Open rate

[View Emails](#) · [Download](#) · [Print](#) · [Share](#)

| | | | |
|------------------------------|-------|------------------------------|-------|
| Overall open rate | 42.7% | Overall click rate | 21.3% |
| List average | 45.3% | List average | 20.4% |
| Industry average (Ecommerce) | 27.1% | Industry average (Ecommerce) | 13.7% |

Winning combination



|LNAME|, Fall Shirts Are Here

Sent Aug 06, 2015, 3:00 pm from Shelby (freddie@freddiesjokes.com)

"Testing the order in which the content appears: main alpha image (blue)"

5,000

Sends

45.7%

Open rate

22.8%

Click rate

View Report

Combination results

| Subject | From name | Content description | Sends | Open rate | Click rate |
|---------|-----------|---------------------|-------|-----------|------------|
|---------|-----------|---------------------|-------|-----------|------------|

Treatment Ramp Up

- Vary the amount of people you expose to different version of the product
- Very good approach to scaling your systems
- Automation and optimization of features is less costly
- Must be supported by good randomization algorithms
- No bias in any particular variant of the experiment

Auto abort

- Reduces the percentage of users assigned to the underperforming Treatment to zero.
- Test on more people
- Reduces risk of exposing more users to some error
- "Move fast and break things"



Ramp Up and Auto Abort

When running an A/B test, what fraction of users should be in the manipulation condition at the start? What fraction should you ramp up to? Why does the author recommend doing this?

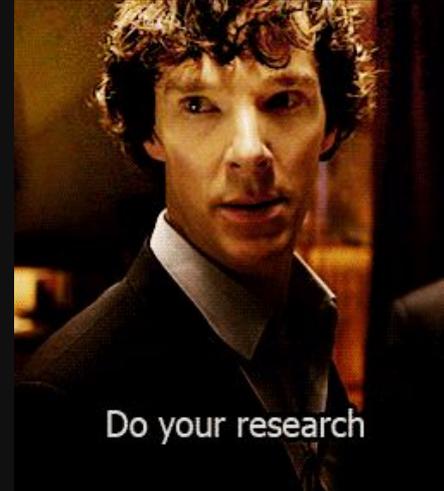
That's what you said...

“The increase ratio sounds rather arbitrary to me, but the benefit of ramping-up in such a fashion is that egregious errors can be caught early-on; because the sensitivity is squared in the formula for sample size, when it is relaxed to, say, supporting detecting a 20% change instead of 5% (thereby detecting bigger changes in behavior, which are most likely caused by egregious errors), the sample size needed decreases by a factor of $4^2 = 16$.”

- Hyeonsu

Discuss with your research partner!

How will you apply A/B testing to your research project? Discuss with your partner for 1 minute.





Academia vs Industry Perspectives

Experimental evidence of massive-scale emotional contagion through social networks

Kohavi, Longbotham, Sommerfield, Henne

There are criticisms of existing methods.

- Social Interaction vs Emotion**
- Non Verbal vs Verbal cues**

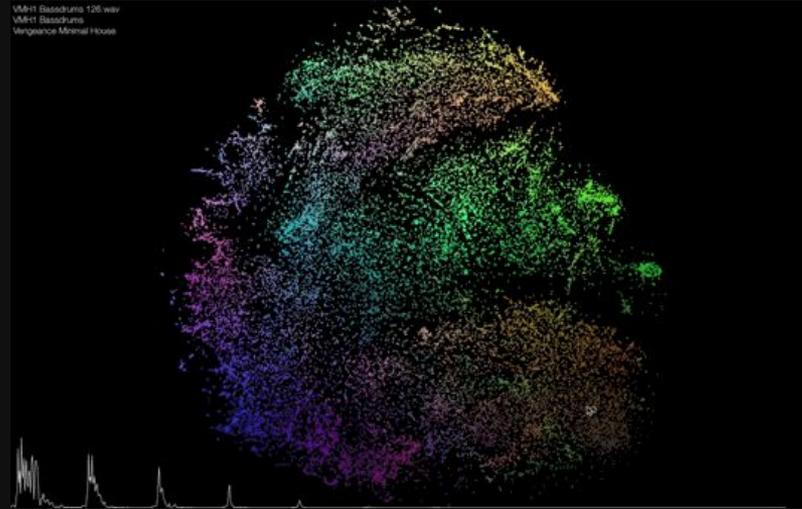
- **Facebook undertook a massive study (~700,000 people) with Cornell University to answer some of these questions.**
- **They wanted to know if people's exposure to emotions on their news feed affected their emotion and thereby the content they posted. How did it affect it?**

Two Parallel Experiments:

- **Exposure to friends' positive emotional content was reduced.**
- **Exposure to friends' negative emotional content was reduced.**
- **There was also a control condition in which a similar proportion of posts in their news feed was omitted at random"**

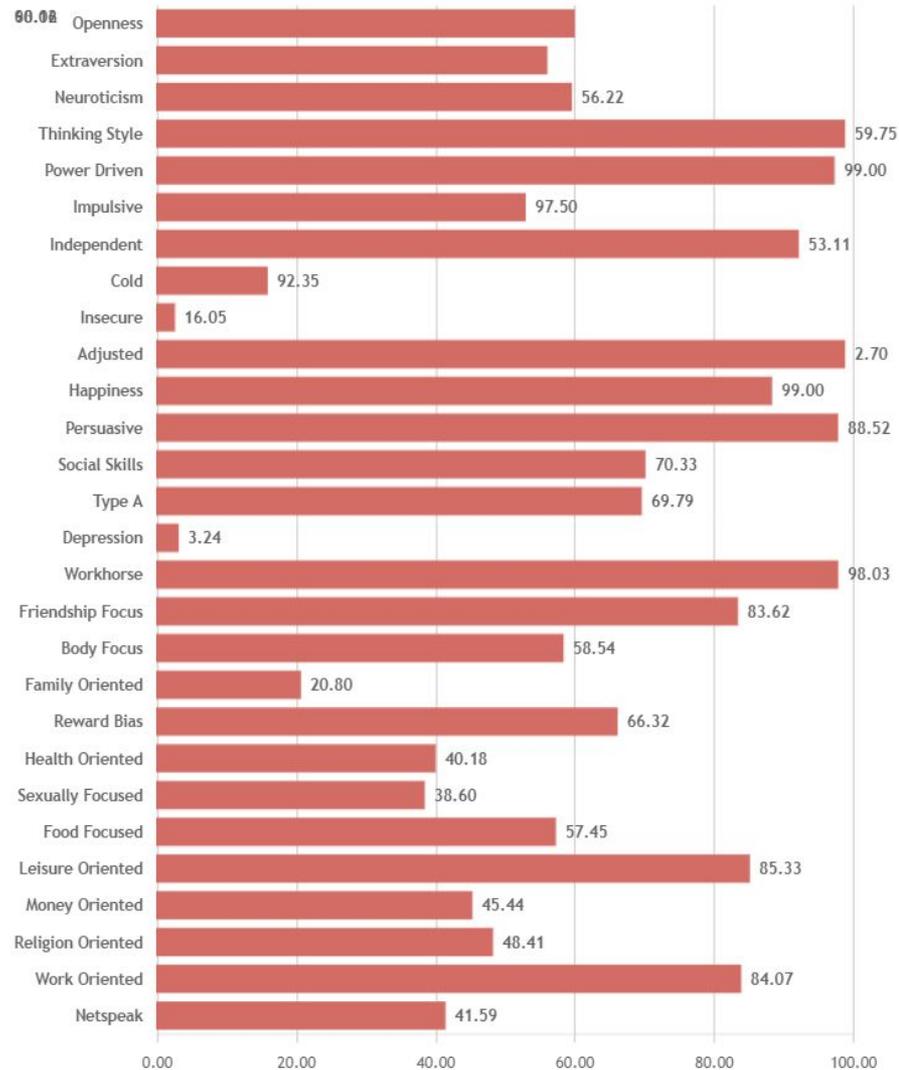
Method

Posts were determined as positive or negative if they contained at least ONE positive or negative word as defined by the Linguistic Inquiry and Word Count (LIWC) software.



LIWC is a Natural language processing tool.

Sentiment analysis of a Twitter account using LIWC



Sentiment analysis of Scott's Twitter account using LIWC



Scott Klemmer

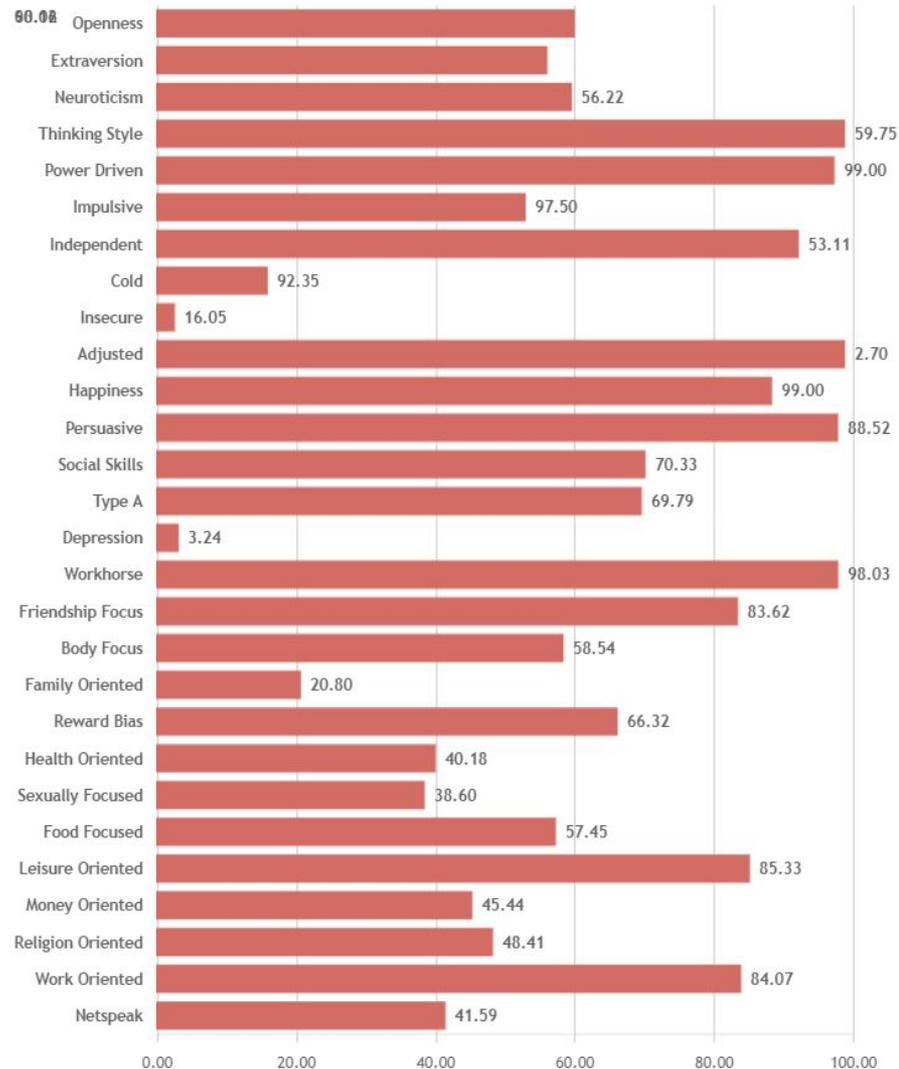
@DesignAtLarge

Associate Professor @DesignLabUCSD;
Coursera #IxDOnline.

📍 La Jolla, San Diego

🔗 d.ucsd.edu/srk

📅 Joined August 2013



They measured two variables:

- **Percentage of all words produced by person that were positive**
- **Percentage of all words produced by person that were negative**

Results

1) Positive Posts Reduced Condition:

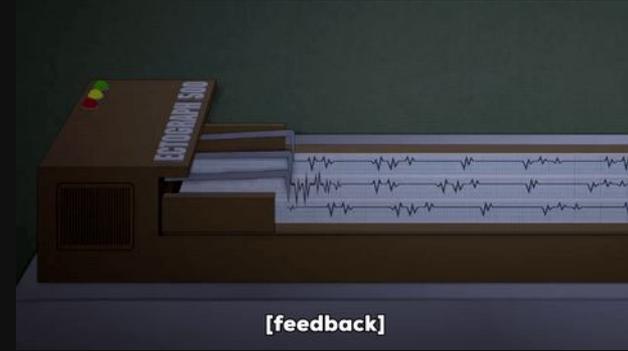
(a) % of positive words decreased by 0.1%

(b) % of negative words increased by 0.04%

2) Negative Posts Reduced Condition:

(a) % of negative words decreased by 0.07%

(b) % of positive words increased by 0.06%



Results proved emotional contagion exists. Emotions expressed by our friends are transferred over to us over large scale social networks. Also proved that nonverbal behavior is not necessary for emotional contagion.

Group Discussion



Turns out people didn't care about the results as much as the ethical dilemma surrounding the study. Discuss with your partner for 1 minute about how ethical/unethical this was. If it was unethical, what part of this process would you mitigate for it to be ethical?

