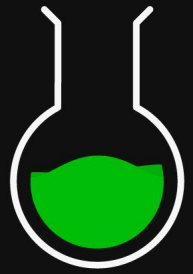


Web Experiments

Neil Sengupta and Calvin Xavier Gomez

Learning Goals 🤘

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

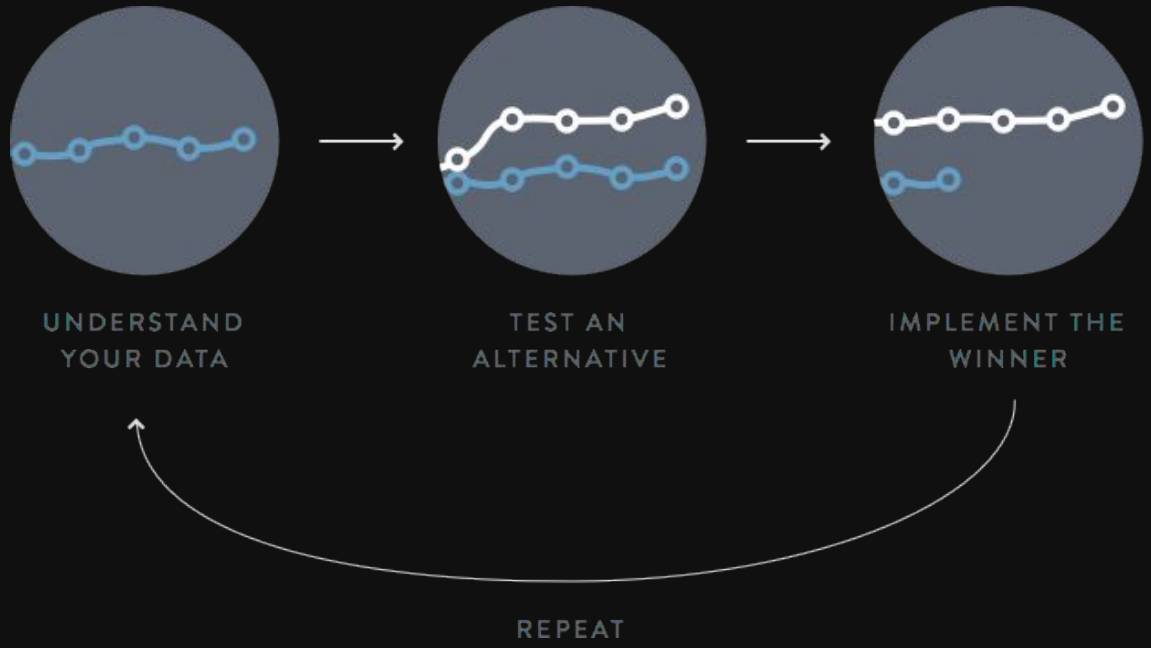
Controlled experiments on the web

Kohavi, Longbotham, Sommerfield, Henne

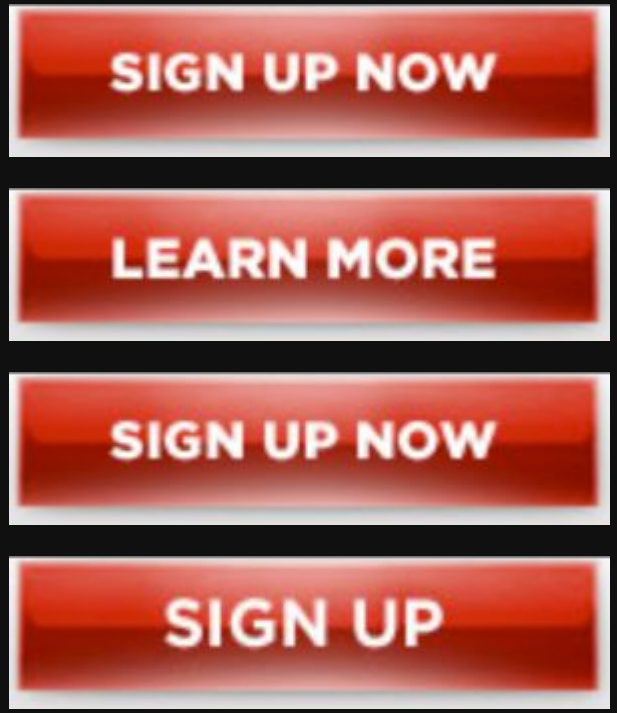
Key Ideas to take away from this paper

- 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

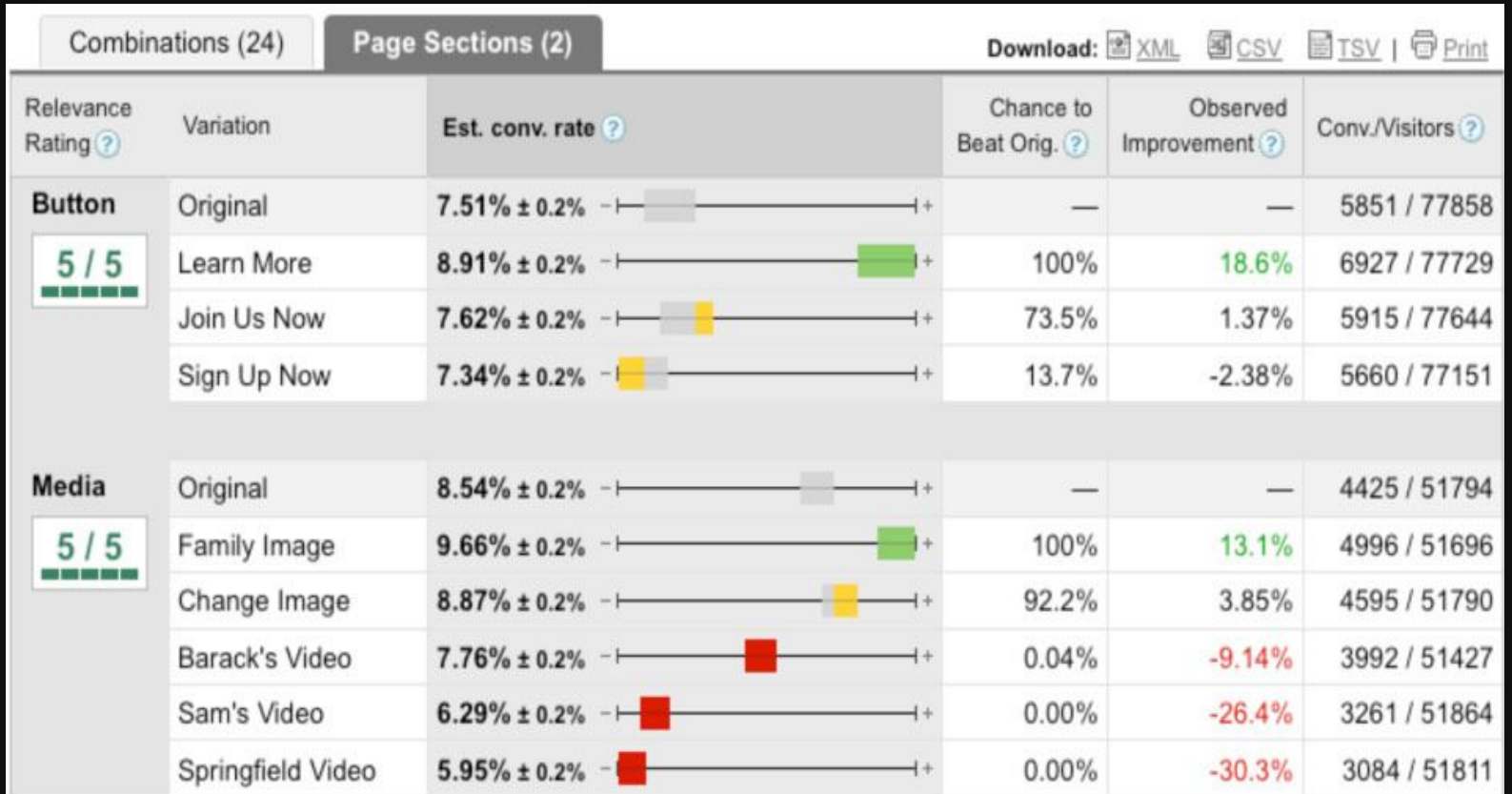


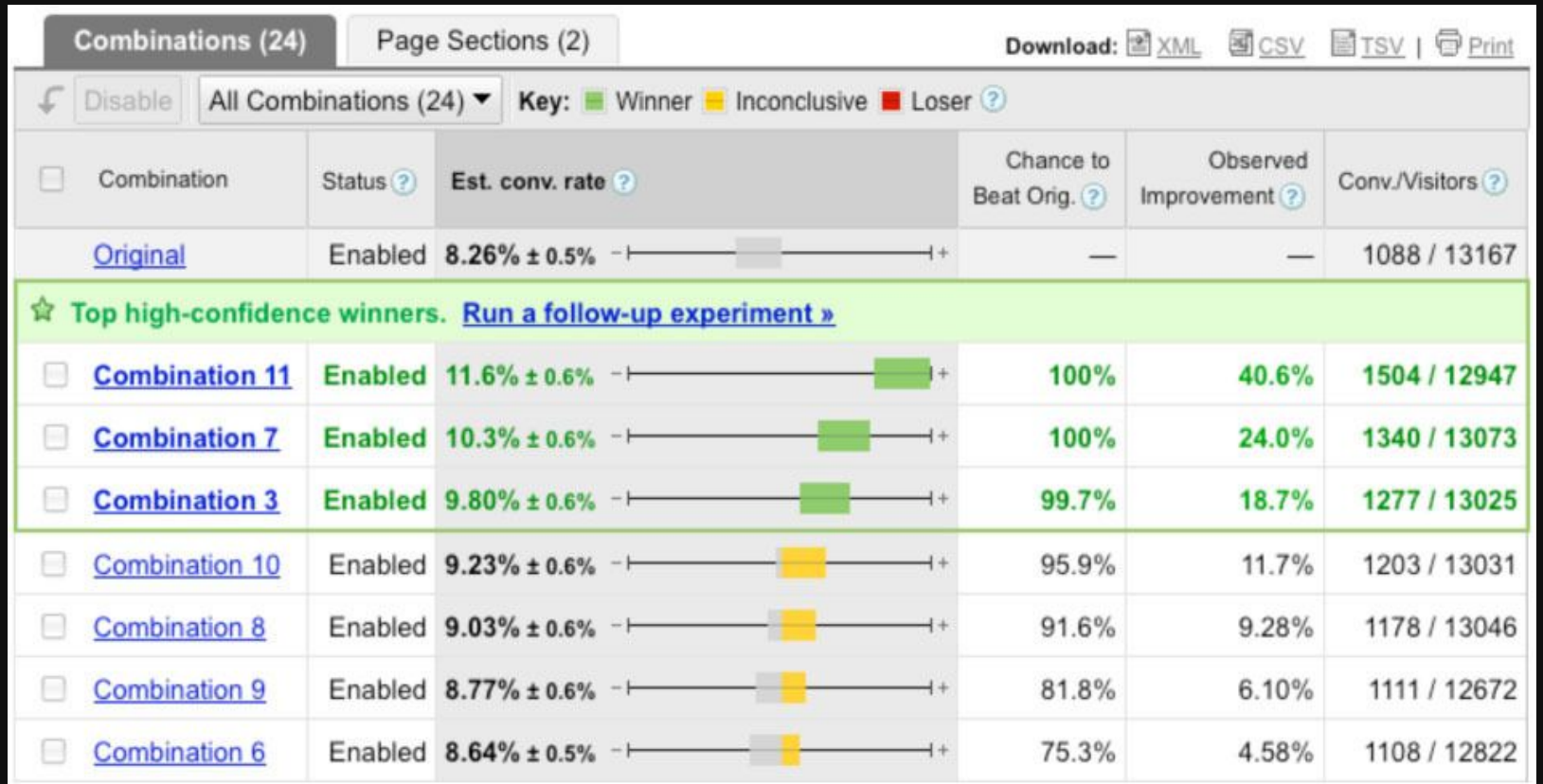
Thanks Obama!


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
5 / 5	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%**	—	—	4425 / 51794
5 / 5	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

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 »					
Combination 11	Enabled	11.6% ± 0.6%	100%	40.6%	1504 / 12947
Combination 7	Enabled	10.3% ± 0.6%	100%	24.0%	1340 / 13073
Combination 3	Enabled	9.80% ± 0.6%	99.7%	18.7%	1277 / 13025
Combination 10	Enabled	9.23% ± 0.6%	95.9%	11.7%	1203 / 13031
Combination 8	Enabled	9.03% ± 0.6%	91.6%	9.28%	1178 / 13046
Combination 9	Enabled	8.77% ± 0.6%	81.8%	6.10%	1111 / 12672
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

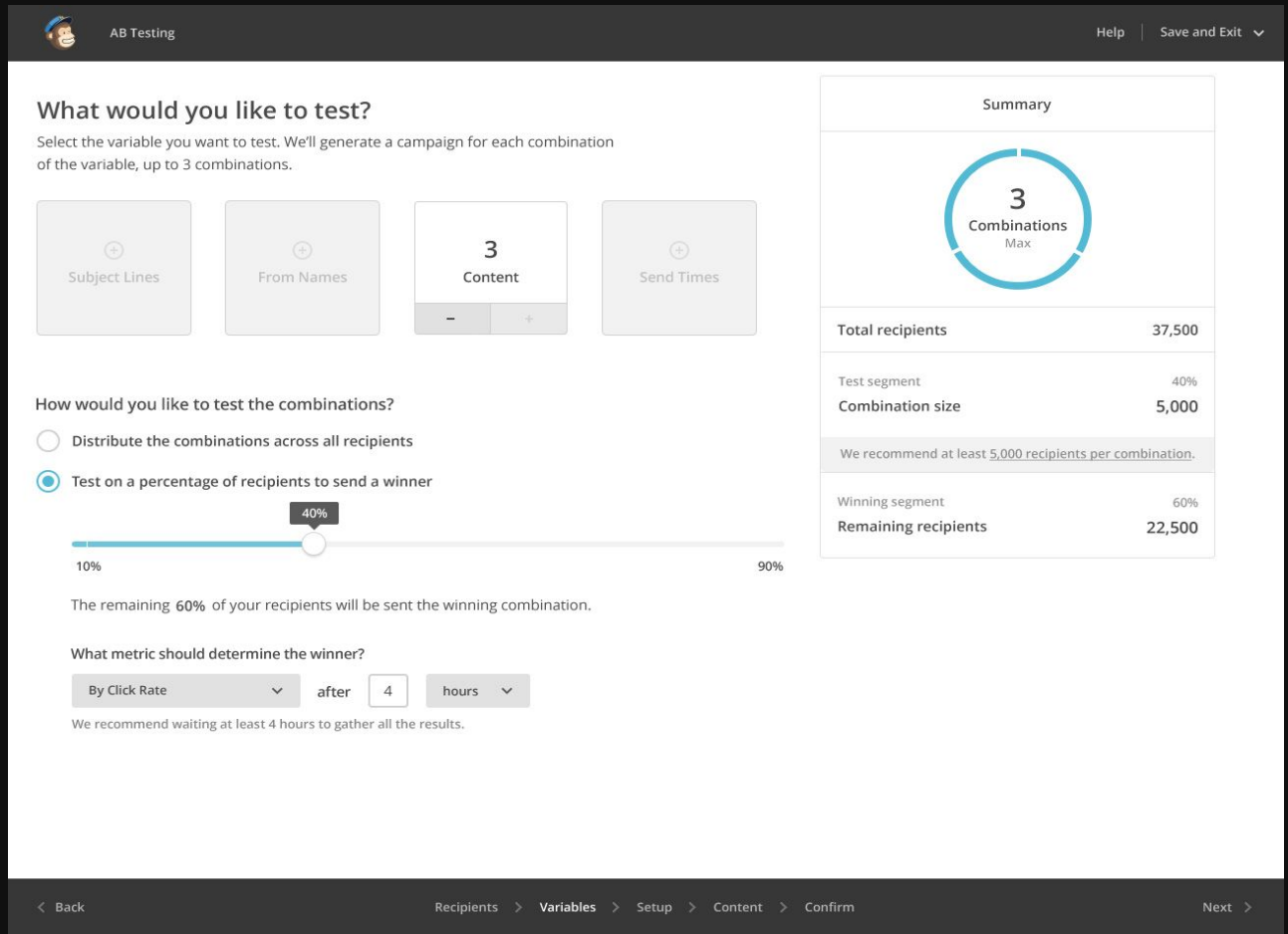
Zip Code

LEARN MORE

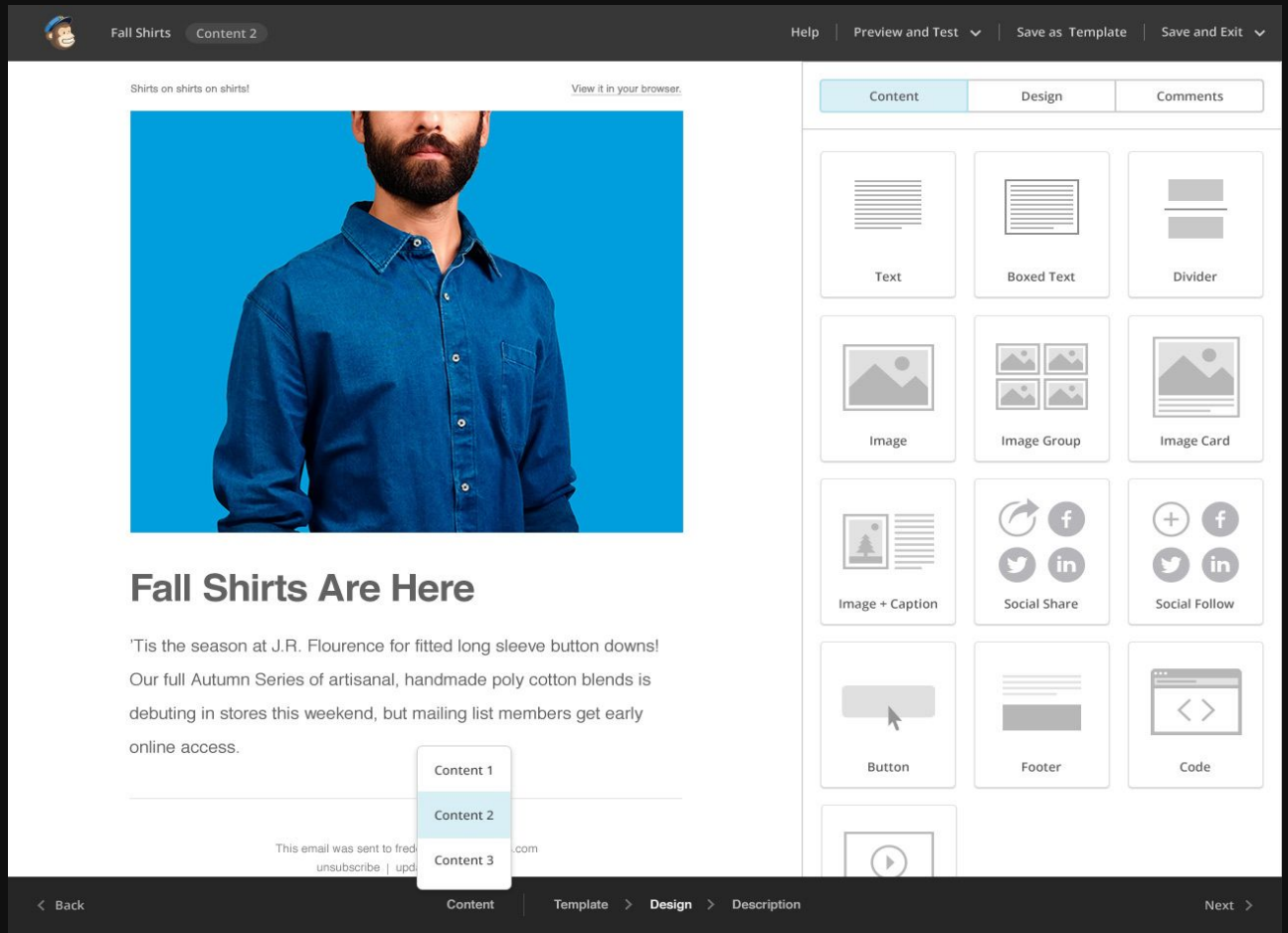
PAID FOR BY OBAMA FOR AMERICA

CONTINUE to WEBSITE

Mailchimp's A/B Test



Mailchimp's A/B Test



Mailchimp's A/B Test

Fall Shirts

Help Save 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

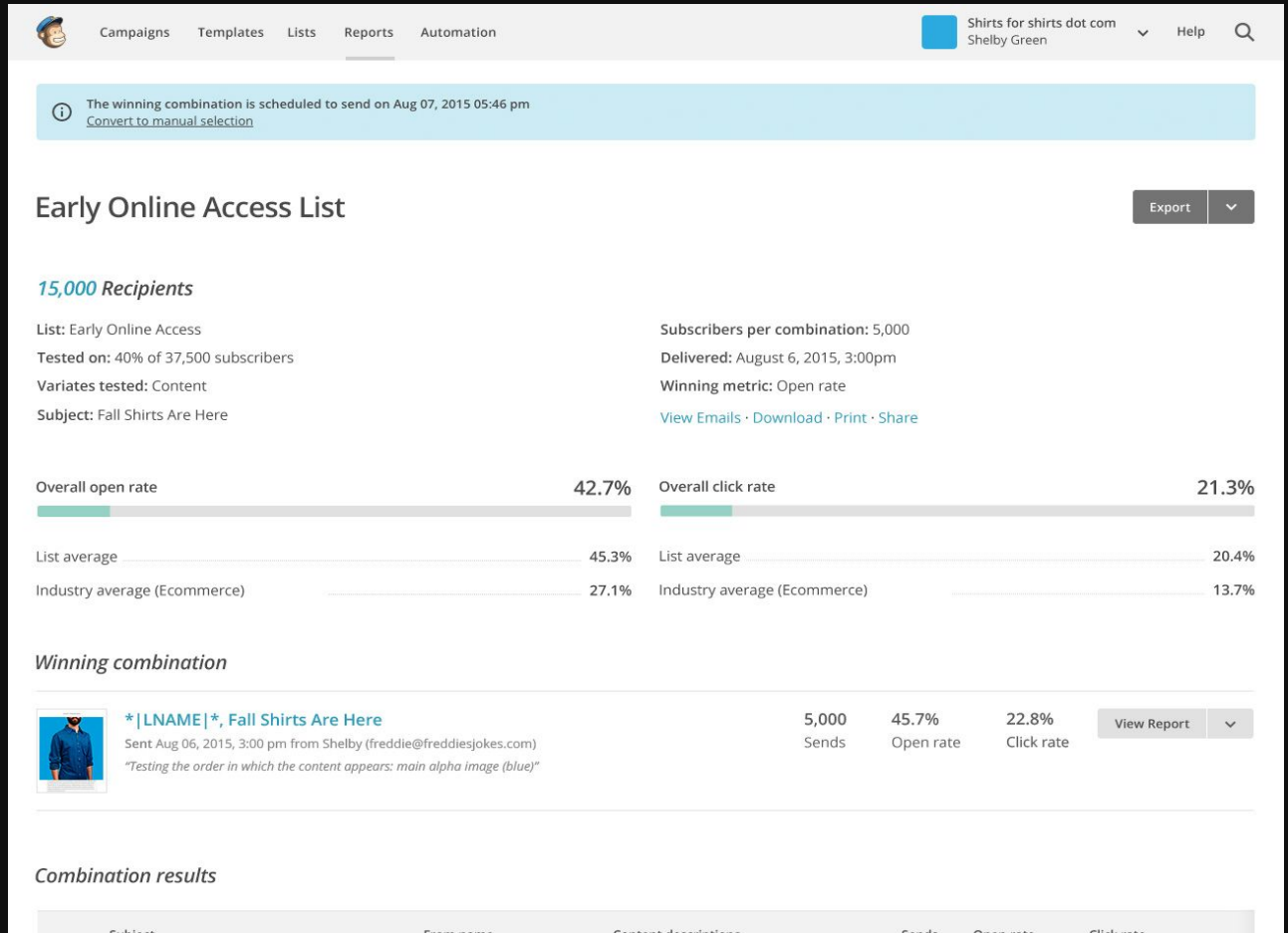
"Testing the order in which the content appears: main alpha image (yellow), headline,"

Back

Recipients > Variables > Setup > **Content** > Confirm

Next

Mailchimp's A/B Test



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.

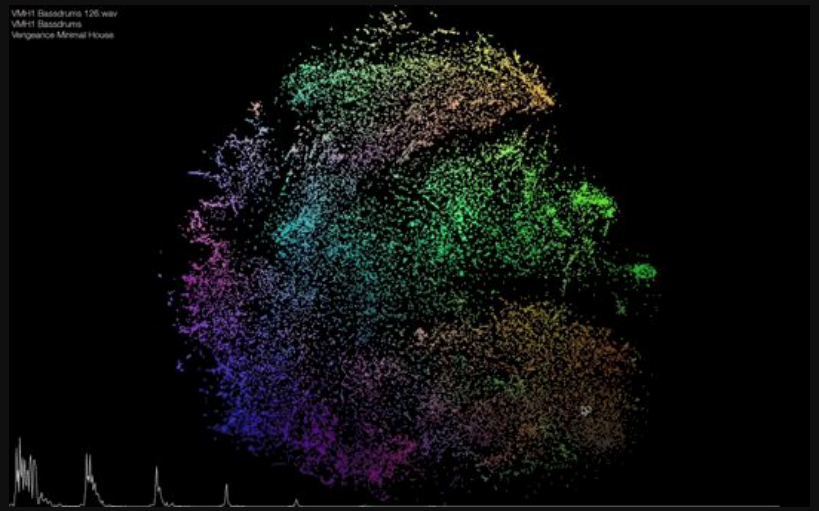
- They wanted to know if people's exposure to emotions on their news feed affected their emotion and thereby the content they posted. How did it affect it?

Method

Two Parallel Experiments:

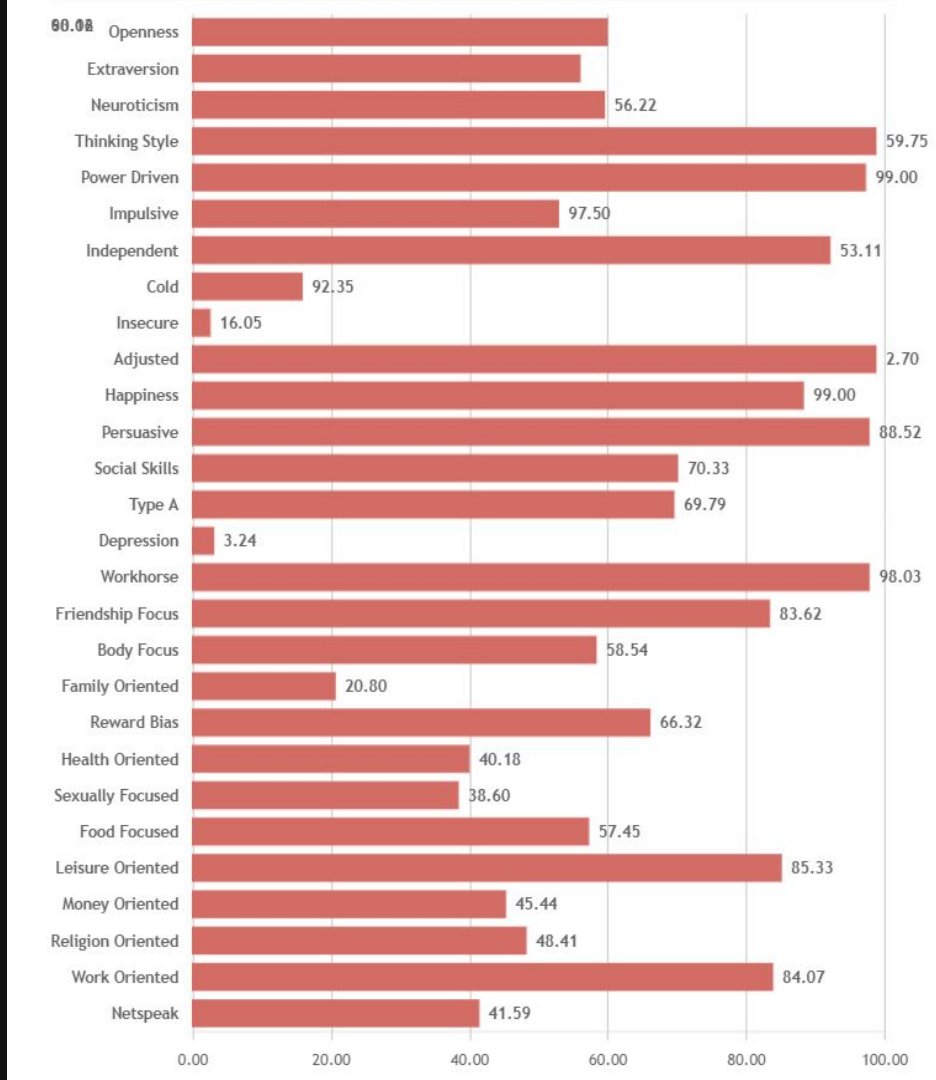
- Exposure to friends' positive emotional content was reduced.
- Exposure to friends' negative emotional content was reduced.
- There was also a control condition in which a similar proportion of posts in their news feed was omitted at random"

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.

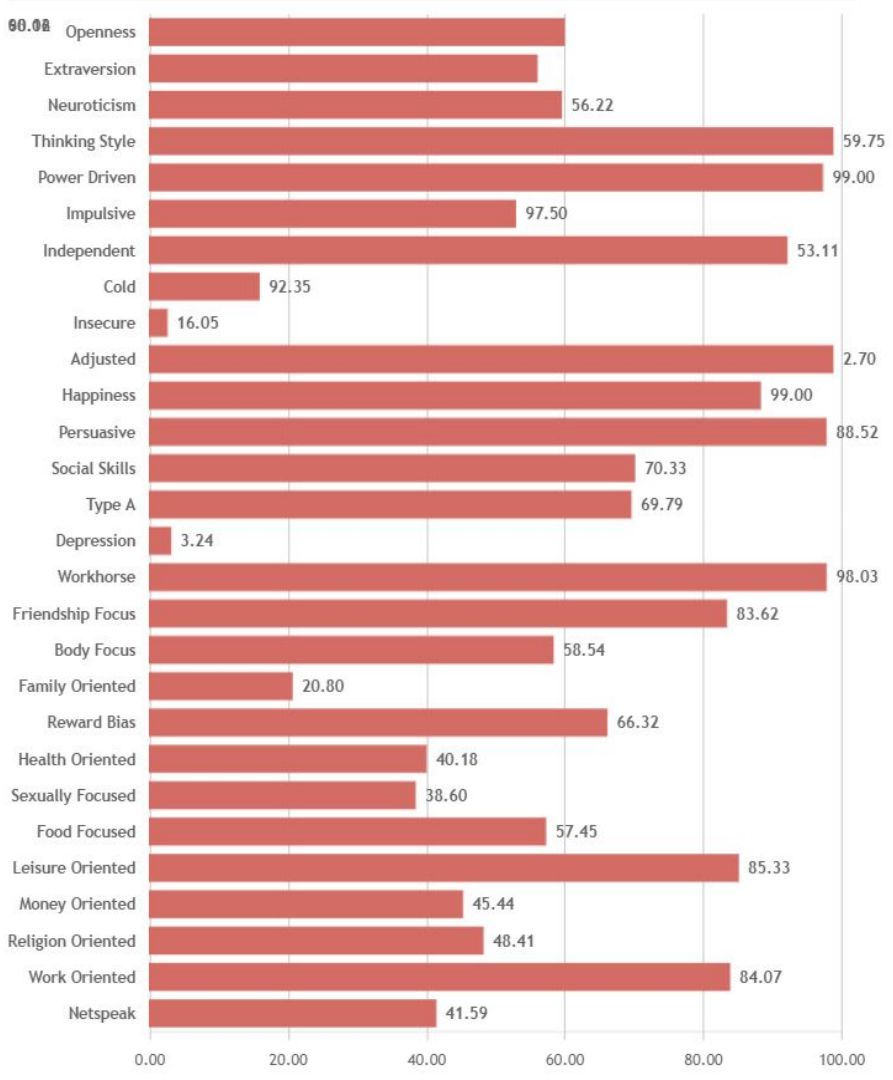


LIWC is a Natural language processing tool.

Sentiment analysis of a Twitter account using LIWC



Sentiment analysis of Scott's Twitter account using LIWC



They measured two variables:

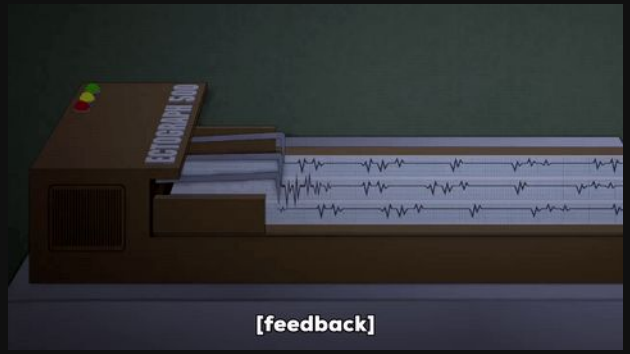
- Percentage of all words produced by person that were positive

- 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?