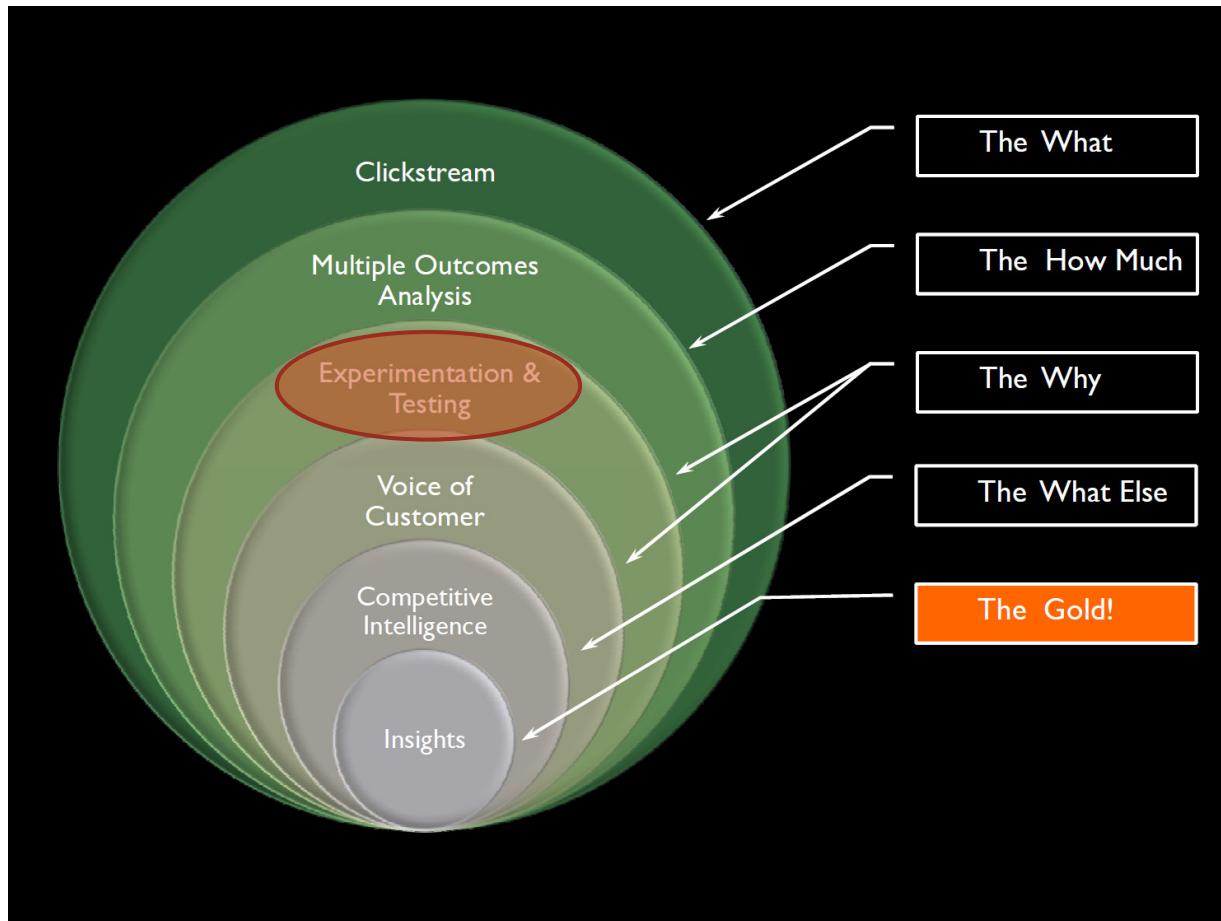


DIGITAL EXPERIMENTS

PROF. XINXIN LI



LEVERAGING THE POWER OF EXPERIMENTATION AND TESTING TOOLS



We should use the A/B testing methodology a LOT more than we do today

-- Bill Gates. 2008

From theory to practice...

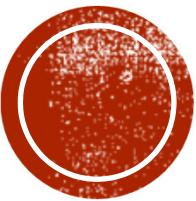
“The difference between theory and practice is greater in practice than in theory”



Learning from Microsoft’s experience --



“Trustworthy Online Controlled Experiments: Five Puzzling Outcomes Explained” Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, Ya Xu.

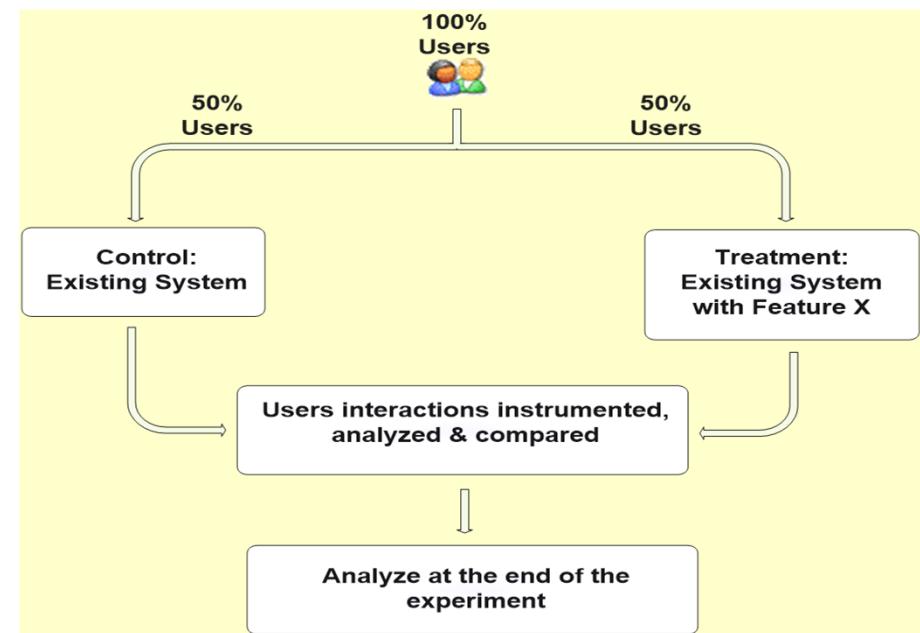


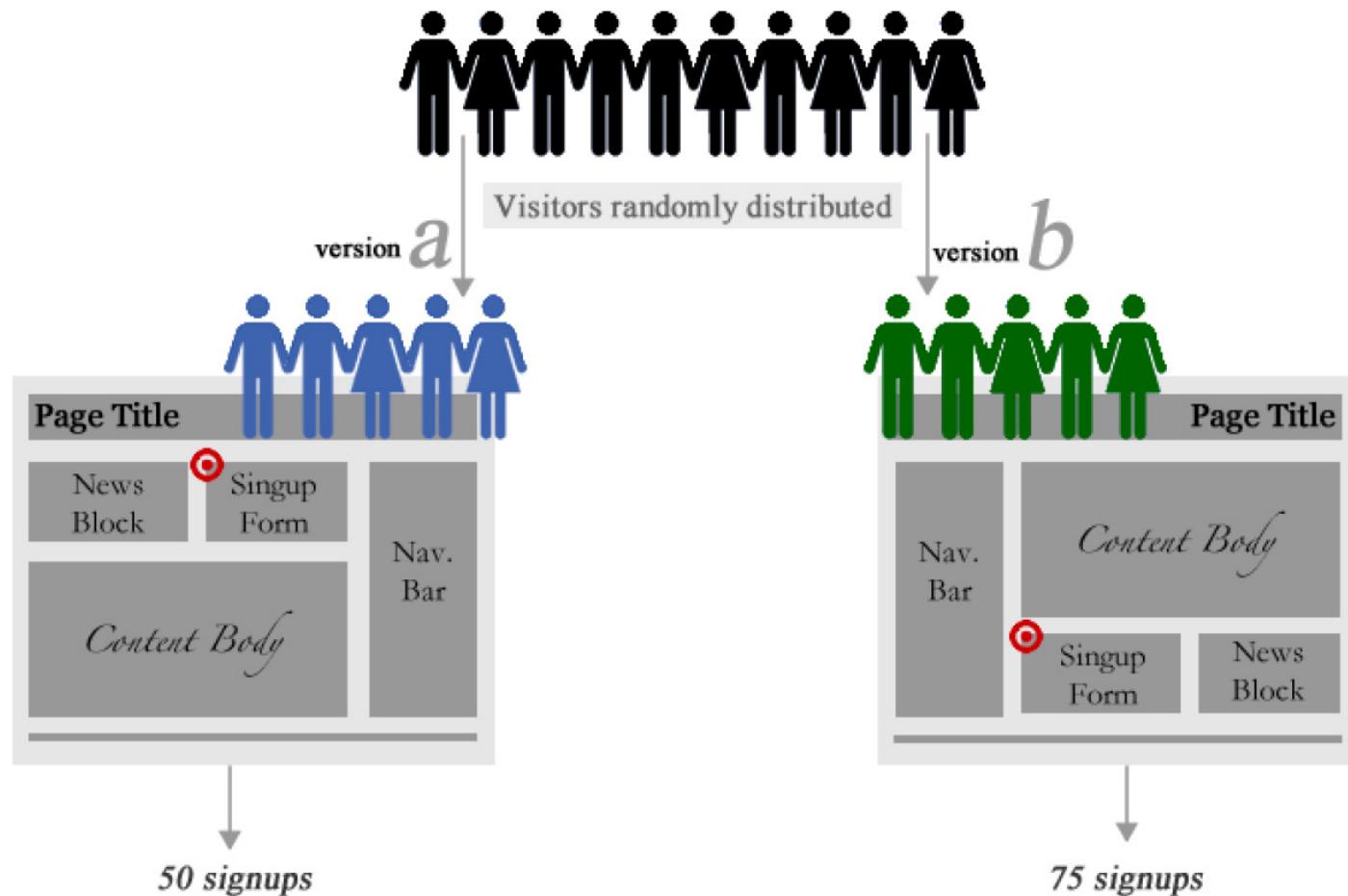
I. CONTROLLED EXPERIMENTS



CONTROLLED EXPERIMENTS

- Concept is not hard
 - A/B testing (field experimentation in the naturally occurring environment of the digital world) **randomly** split traffic between two (or more) versions
 - A (Control)
 - B (Treatment)
 - Collect metrics of interest
 - Analyze
- Need to run statistical test to confirm differences are not due to chance
- A scientific way to prove **causality**, i.e., the changes in metrics are caused by changes introduced in the treatment





EXPERIMENTS CAN HELP COMPARE...

- Bid changes for ad-groups
- Bid changes for individual KWs
- Effect of New Ad-groups
- New Keywords
- Different KW match types
- Different ad variations
- Different landing page design
- etc.

GOALS OF EXPERIMENTS

- Testing of frontend sites
 - e.g., which design for the “Find a home” widget on MSN Real Estate site is better?

Find a new home or apartment

Existing Homes from REALTOR.com®
 Foreclosures from RealtyTrac.com™

Price Range: \$0 — No Maximum
Enter City Select a State Or Enter ZIP Go

Senior Living • Home Plans Control



Existing Homes Foreclosures New Construction Rentals

Find Existing Homes for Sale

Enter City State
OR
Enter Zip Find homes ▶

Treatment 1



Existing Homes Foreclosures New Construction Rentals

Find Existing Homes for Sale

Enter City State
or
Enter Zip Find homes ▶

Treatment 2



What are you looking for?

Existing Homes Enter City State
New Construction Enter Zip
Rentals
Foreclosures
Senior Living
Home Valuation
Professional Services
Find homes ▶

Treatment 3



Find a new Home or Apartment

Existing Homes FOR SALE New Construction Foreclosures FOR RENT Rentals

Enter Zip or Enter City State ▾ Search listings ▶

Treatment 4



Find Your Dream Home or Apartment

City, State or ZIP

Existing homes New construction
 Foreclosures Rentals

Search listings ▶

Treatment 5



- Testing of backend algorithms
 - E.g., when someone searches for “Chess” should an informative website like Wikipedia entry on chess come up first, or should sites providing the game online come up first?

EXAMPLE: CHECKOUT PAGE FOR DOCTOR FOOTCARE

- Which one has a higher conversion rate?

Doctor FootCare™

[Home](#) | [Products](#) | [Learn More](#) | [Tips](#) | [Testimonials](#) | [FAQ](#) | [About Us](#) | [Contact Us](#) 1-866-211-9733

Shop With Confidence

Satisfaction Guaranteed 30-day, hassle-free Returns
 100% Safe, **Secured** shopping We assure your Privacy

100% Secured Checkout

Item Name	Item Number	Quantity	Remove	Unit Price	Subtotal
Trial Kit	FFCS	<input type="text" value="1"/>		\$0.00	\$0.00
			<input type="button" value="Update"/>	Total: \$0.00	

Select Shipping Method Standard (\$5.95)

100% Secured Checkout

[Continue Shopping](#) [> Proceed To Checkout](#)

[Home](#) | [Products](#) | [Learn More](#) | [Tips](#) | [Testimonials](#) | [FAQ](#) | [About Us](#) | [Contact Us](#) | [Shopping Cart](#)

Doctor FootCare™

[Home](#) | [Products](#) | [Learn More](#) | [Tips](#) | [Testimonials](#) | [FAQ](#) | [About Us](#) | [Contact Us](#) 1-866-211-9733

Shop With Confidence

Satisfaction Guaranteed 30-day, hassle-free Returns
 100% Safe, **Secured** shopping We assure your Privacy

100% Secured Checkout

Item Name	Item Number	Quantity	Remove	Unit Price	Subtotal
Trial Kit	FFCS	<input type="text" value="1"/>		\$0.00	\$0.00
			<input type="button" value="Update"/>	<input type="button" value="Discount"/>	<input type="button" value="Total"/>

Enter Coupon Code

Select Shipping Method Standard (\$5.95)

100% Secured Checkout

[Recalculate](#) [Continue Shopping](#) [> Proceed To Checkout](#)

[Home](#) | [Products](#) | [Learn More](#) | [Tips](#) | [Testimonials](#) | [FAQ](#) | [About Us](#) | [Contact Us](#) | [Shopping Cart](#)

EXAMPLE: FONT COLOR EXPERIMENT

- Can you tell the difference

WEB IMAGES VIDEOS MAPS NEWS MORE

bingamazon

ALSO TRY: Walmart · Target · Netflix

202,000,000 RESULTS Any time ▾

Ad related to amazon

[Amazon.com® Official Site - Huge Selection and Amazing Prices.](#)
www.Amazon.com · Official Site
Free Shipping on Orders Over \$25

Amazon Prime Prime Instant Video
Join Amazon Student Rent Textbooks
Download eTextbooks Kindle Store

[Amazon.com: Online Shopping for Electronics, Apparel, Computers ...](#)
www.amazon.com ▾ Official site
Customer service 866-216-1072
Online shopping from the earth's biggest selection of books, magazines, music, DVDs, videos, electronics, computers, software, apparel & accessories, shoes, jewelry ...

Books
Shop online for millions of new and used books

Music
Browse best sellers, new releases, deals, vinyl records, MP3s, and ...

WEB IMAGES VIDEOS MAPS NEWS MORE

bingamazon

ALSO TRY: Walmart · Target · Netflix

202,000,000 RESULTS Any time ▾

Ad related to amazon

[Amazon.com® Official Site - Huge Selection and Amazing Prices.](#)
www.Amazon.com · Official Site
Free Shipping on Orders Over \$25

Amazon Prime Prime Instant Video
Join Amazon Student Rent Textbooks
Download eTextbooks Kindle Store

[Amazon.com: Online Shopping for Electronics, Apparel, Computers ...](#)
www.amazon.com ▾ Official site
Customer service 866-216-1072
Online shopping from the earth's biggest selection of books, magazines, music, DVDs, videos, electronics, computers, software, apparel & accessories, shoes, jewelry ...

Books
Shop online for millions of new and used books

Music
Browse best sellers, new releases, deals, vinyl records, MP3s, and ...



WHAT ABOUT OBSERVATIONAL STUDIES?

- Why not analyze using historical data?
- Here's an example of Kindle Sales over time.
The site was changed, and there was an amazing spike



EXTERNAL EVENTS CAN DWARF YOUR CHANGES



- In this example of an A/B test, you'd be better off with version A
- In controlled experiments, both versions are impacted the same way by external events

EXAMPLE: CORRELATIONS DUE TO COMMON CAUSE

- Article in Nature: leaving the night-light on in the nursery made the child myopic later in life
- CNN publishes summary of article
- A year later, an important observation was made:
 - Myopic parents are more likely to leave the light on, and
 - Myopic parents are more likely to have myopic childrenA common factor was identified
- Once you control for the parents' myopia, the relationship in the original study disappears
- See [How Science Works: The Story of Night-Light Myopia](#)

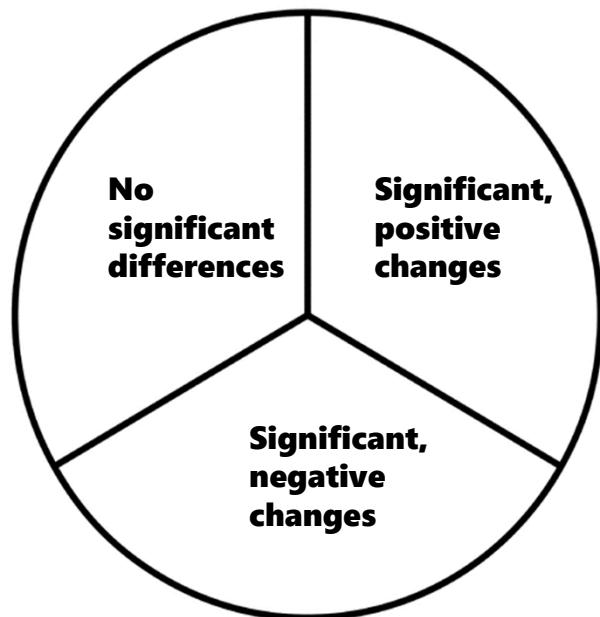
ADVANTAGE OF CONTROLLED EXPERIMENTS

- Controlled experiments test for **causal** relationships, not simply correlations
- When the variants run concurrently, only two things could explain a change in metrics:
 1. The “feature(s)” (A vs. B)
 2. Random chance

Everything else happening affects both variants
For #2, we conduct statistical tests for significance
- The gold standard in science and the only way to prove efficacy of drugs in FDA drug tests
- Controlled experiments are not the panacea for everything
 - Primacy and newness effects
 - Short term vs. long term effects
 - Show result, but not explanation
 - Spillover effects
 - etc.

OUR INTUITION IS POOR

- Ideas tested at Microsoft



Google ran approximately 12,000 randomized experiments in 2009, with [only] about 10 percent of these landing to business changes."

--*Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics and Society* (Jim Manzi)

80% of the time you/we are wrong about what a customer wants.

--*Web Analytics* (Avinash Kaushik 2006)

ORGANIZATIONAL EXPERIMENTATION SYSTEM

- Conduct a lot of experiments, or, conduct few carefully selected experiments
- Cost of experimentation can be high. So, consider a lot of ideas but narrow down to fewer. Use aggregate metrics (e.g. Google Analytics) to help focus on useful hypothesis worth testing
- Use qualitative feedback (e.g. surveys, usability studies in lab) etc. to complement studies from experiments



II. EXPERIMENTAL DESIGN



DECIDING THE OUTCOME VARIABLE

- Important to find metrics that are measurable over the short term but are predictive of long term goals
- E.g., search engines consider the number of queries per user and revenue per user as important long term goals.
- Bing had a bug in an experiment which degraded search results. Yet number of queries per user and revenue per user increased in the short term. Why?

DECIDING THE OUTCOME VARIABLE

- Important to find metrics that are measurable over the short term but are predictive of long term goals
- E.g., search engines consider the number of queries per user and revenue per user as important long term goals.
- Bing had a bug in an experiment which degraded search results. Yet number of queries per user and revenue per user increased in the short term. Why?
 - Degradation of SE algorithm forces people to issue more queries and click on more ads as they cannot easily find what they are looking for
 - So, there would be a short term increase in these metrics until users switch to alternatives leading to a long term decline
- What outcome variable to choose then?

DECIDING THE OUTCOME VARIABLE

- Queries per user in a certain time period can be decomposed as:

$$\frac{\text{sessions}}{\text{user}} \times \frac{\text{queries}}{\text{session}}$$

DECIDING THE OUTCOME VARIABLE

- Queries per user in a certain time period can be decomposed as:

$$\frac{\text{sessions}}{\text{user}} \times \frac{\text{queries}}{\text{session}}$$

- Queries/session is a measure of queries/task (which should be minimized subject to the task being successfully completed). An increase in this metric means that algorithm has degraded, but a decrease may also mean that users are abandoning use. So, direction not clearly related to outcomes

DECIDING THE OUTCOME VARIABLE

- Queries per user in a certain time period can be decomposed as:

$$\frac{\text{sessions}}{\text{user}} \times \frac{\text{queries}}{\text{session}}$$

- Queries/session is a measure of queries/task (which should be minimized subject to the task being successfully completed). An increase in this metric means that algorithm has degraded, but a decrease may also mean that users are abandoning use. So, direction not clearly related to outcomes
- Sessions/user – if there are more sessions per users, they are likely using the SE for more search tasks. The direction is clear here → good outcome measure to use

DECIDING THE OUTCOME VARIABLE

- Likewise, when looking at revenue metrics, we want to increase revenue per user without negatively impacting engagements metrics like sessions/user or the number of users
- An example
 - At Amazon, outcome of different types of email promotions was evaluated on purchases whose sessions were referred by emails
 - Tests showed significant increase in revenue per user for almost all types of promotions; and implemented; but complaints came in
 - What could be wrong?

DECIDING THE OUTCOME VARIABLE

- Likewise, when looking at revenue metrics, we want to increase revenue per user without negatively impacting engagements metrics like sessions/user or the number of users
- An example
 - At Amazon, outcome of different types of email promotions was evaluated on purchases whose sessions were referred by emails
 - Tests showed significant increase in revenue per user for almost all types of promotions; and implemented; but complaints came in
 - What could be wrong?
- Many customers unsubscribed and could no longer be targeted in future. We can add a penalty as the number of unsubscribes times lifetime value from a customer. This updated metric caused many email campaigns to evaluate negative

DECIDING THE OUTCOME VARIABLE

- Find metric with clear directional impact on long-term business outcomes. **Focus on long-term lifetime customer value**
- Improving some feature specific metric may be easy, but better to have a comprehensive metric such as sessions/user, or time to success
- It is easy to improve metrics for a small segment, e.g. key metrics for weather related queries on Bing can be improved relatively easily; but a 10% improvement to a 1% segment may only impact the average by approximately 0.1%

HOW LONG SHOULD AN EXPERIMENT BE RUN?

- Related to the requirement of sample size
- When “novelty” or “primacy” effects are suspected (i.e. initial effect of experiment is not the same as the long-term effect), experiment should be run long enough to get the asymptotic effect of experiment
- Choice of outcome metric has a role
 - Some metrics have higher/lower variance by their nature, e.g. conversion probability has lower variance than number of purchase units (small integers), which has lower variance than revenue (real-valued)
 - Bigger sample (longer experiment) would be required with high variance metrics to get sufficient power of the test
- Ratio of users assigned to control and test groups also matters
 - Equal split requires smaller sample size

TRADITIONAL SAMPLE SIZE DETERMINATION

- Choose sample size for 80% power

$$n_1 = n_2 \approx \left(Z_{\frac{1-\alpha}{2}} + Z_{\beta} \right)^2 \left(\frac{s_1^2 + s_2^2}{\delta^2} \right)$$

- s_1, s_2 : standard deviation of response
- δ : difference (between null hypothesis and alternative hypothesis) to detect
- α : chance of false positive (type I error when null hypothesis is true)
- β : chance of false negative (type II error when alternative hypothesis is true)

e.g.,

- 95% statistical significance ($\alpha = 0.05$)
- 80% statistical power ($\beta = 0.2$)
- Detectable difference 1% ($\delta = 0.01$)
- Conversion rate 15% ($s_1^2 \approx s_2^2 \approx 0.15(1 - 0.15) = 0.1275$)

$$n_1 = n_2 \approx (1.96 + 0.84)^2 \left(\frac{0.1275 + 0.1275}{0.01^2} \right) \approx 20,000$$

TRADITIONAL SAMPLE SIZE DETERMINATION

The screenshot shows a web-based sample size calculator. At the top, there is a navigation bar with the AB TASTY logo and links for Solutions, Clients, Pricing, Resources, Blog, and Company. Below the navigation, the main heading is "How many users do you need?". To the left, there are four input fields with their respective values and units:

Parameter	Value	Unit
Conversion Rate [?]	15	%
Minimum Detectable Effect [?]	6.67	%
Statistical Significance [?]	95	%
Statistical Power [?]	80	%

To the right of these inputs is a large orange arrow pointing towards a result box. The result box contains the text "Required number of tested visitors per variation" followed by the value "20,156".

<https://www.abtasty.com/sample-size-calculator/>

Detectable difference $\delta = 15\% * 6.67\% \approx 1\%$

PROFIT-MAXIMIZING A/B TESTS

- Balance between having a sufficient sample size for a reliable result and leaving enough people out to benefit from the experiment result
(by Elea Feit and Ron Berman)

$$n_1^* = n_2^* = \sqrt{\frac{N}{4} \left(\frac{s}{\sigma} \right)^2 + \left(\frac{3}{4} \left(\frac{s}{\sigma} \right)^2 \right)^2 - \frac{3}{4} \left(\frac{s}{\sigma} \right)^2}$$

N : is population size

s : standard deviation of response

σ : difference in performance between treatments

e.g. if $N = 10,000$, conversion rate = 15%, difference in performance = 1%

$$\begin{aligned} n_1^* = n_2^* &= \sqrt{\frac{10,000}{4} \left(\frac{\sqrt{.15 * .85}}{.01} \right)^2 + \left(\frac{3}{4} \left(\frac{\sqrt{.15 * .85}}{.01} \right)^2 \right)^2} \\ &\quad - \frac{3}{4} \left(\frac{\sqrt{.15 * .85}}{.01} \right)^2 \approx 1069 \end{aligned}$$

ANALYZING A/B TESTING DATA

- To find the causal effect of treatment on outcome, consider regression

$$\text{Outcome} = a + b * \text{Treatment Dummy}$$

- Outcome in the control group = a
- Outcome in the treatment group = a + b
- If b is significant → Treatment has a significant effect on outcome

ANALYZING A/B TESTING DATA

- What if more than one treatment groups?
 - Treat groups as a categorical variable
 - Create a dummy for each treatment
 - Use the control group as the baseline group

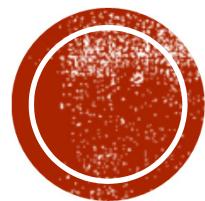
$$\text{Outcome} = a + b * \text{Treatment Dummy 1} + c * \text{Treatment Dummy 2}$$

- Consider heterogeneous treatment effect (i.e., treatment effect differs for different type of users)
 - Use interaction terms (remember the Booksales exercise we did in class?)

$$\begin{aligned}\text{Outcome} = & a + b * \text{Treatment Dummy 1} + c * \text{Treatment Dummy 2} \\ & + d * \text{Differentiator} \\ & + e * \text{Differentiator} * \text{Treatment Dummy 1} \\ & + f * \text{Differentiator} * \text{Treatment Dummy 2}\end{aligned}$$

DEALING WITH ROBOT GENERATED DATA

- For large sites, robots can represent 15-30% of page views. We are not interested in the response of robots to the test. So, data coming from bots must be filtered out before analysis
- Non-malicious robots can be removed relatively easily
 - Robots who do not accept cookies
- For malicious robots (those who accept cookies), use of heuristics would be required e.g.
 - Remove users with an unlikely number of page requests
 - Remove users who issue requests very rapidly



III. PITFALLS TO AVOID WHEN INTERPRETING EXPERIMENT RESULTS



AVOID THE RED HERRINGS

Any piece of data or evidence that looks interesting or different is usually wrong

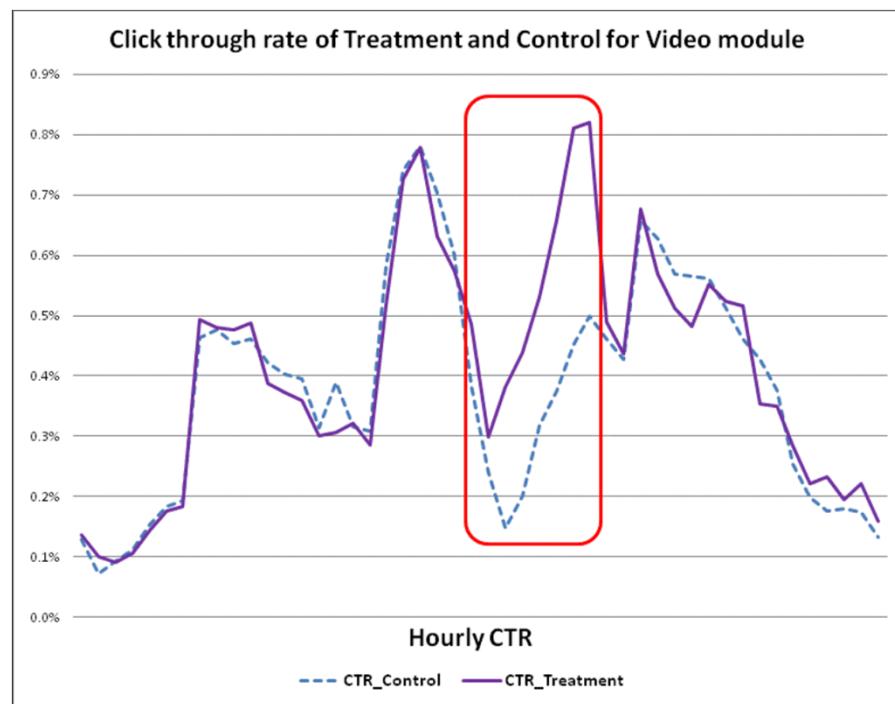
--*Twyman's law*

- For example
 - If you have a mandatory birth date field and people think it's unnecessary, you'll find lots of people born on 11/11/11 or 01/01/01.
 - If you have an optional drop down, do not default to the first alphabetical entry, or you'll have lots of: jobs = Astronaut
 - For most web sites, there is no traffic between 2am and 3am March 12, 2017. Don't worry, there was no outage.

<https://amplitude.com/blog/2017/05/17/twymans-law>

VALIDATION CHECKS: CONSISTENCY OF RESULT

- Suppose the test shows a statistically significant benefit compared to control
- Is the benefit “generalized” over time, or coming from a limited time “big effect”?



Here, the treatment showed a 2% increase over control. But the reason was a 7-hour period where the top headline link was accidentally set differently for control and treatment groups. Removing the 7-hour period from the data showed no significant difference

LESSONS LEARNED

- Experimental control is critical. Keep everything constant except the thing you want to test
- Drill-down by time to look at granular level data
- Use screen scrapers to save screen shots of the pages being experimented on a regular basis

CONSISTENCY OF EXPERIENCE ACROSS TEST AND CONTROL GROUPS IS IMPORTANT

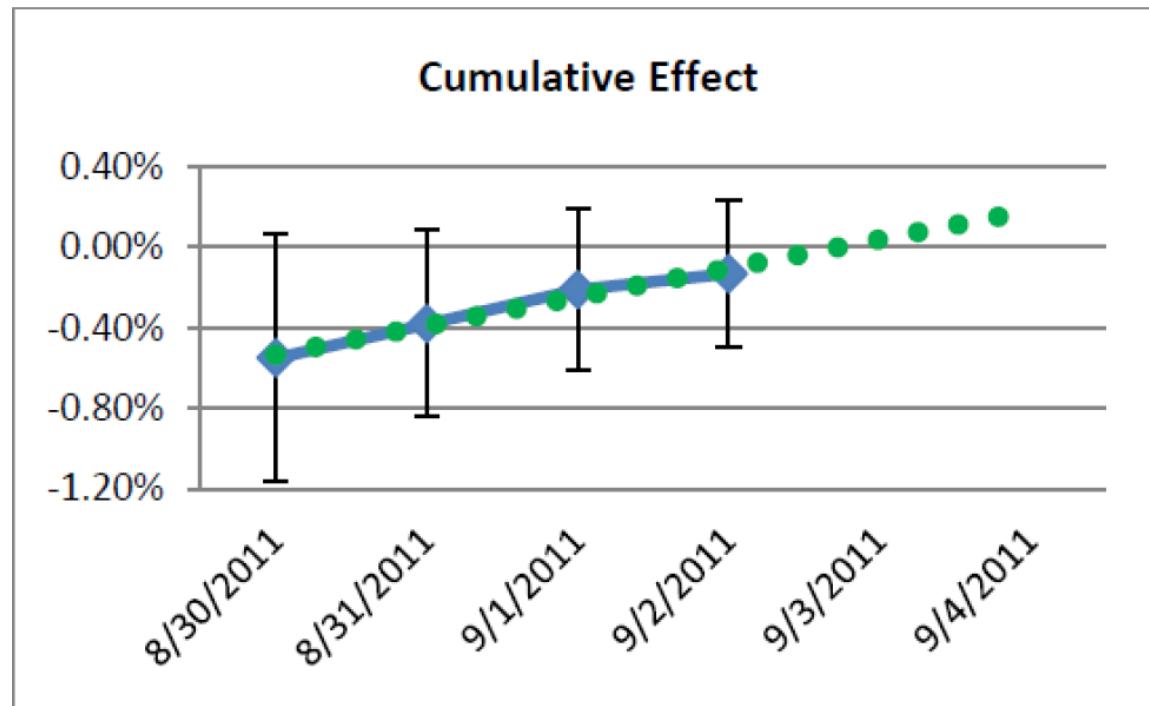
- Effectiveness of the test page in website testing may appear to be less when:
 - using browser redirects to send users to the test page, which may result in higher page load time for the test group
 - Any performance penalty should be equal on test and control groups
 - using servers with different configurations/ capacities for the test and the control pages
 - having differential use of resources on the same server, e.g., if the control group is bigger in size than test group, then more cache capacity may be dedicated to control group pages. This results in quicker loading of control pages compared to test pages.

VALIDATION CHECK: A/A TESTS

- Assign users to two groups and assign exactly the same experience to both
- Run this many times. Using a significance level of 5%, you should not be able to reject the null hypothesis of no difference more than 5% of the time
- Failed A/A tests suggest asymmetric experience between control and treatment groups, e.g.,

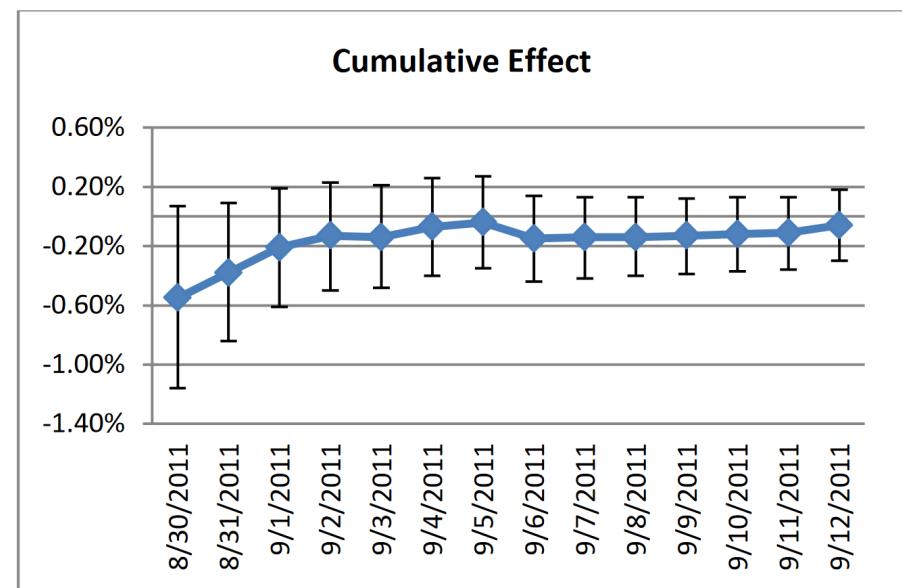
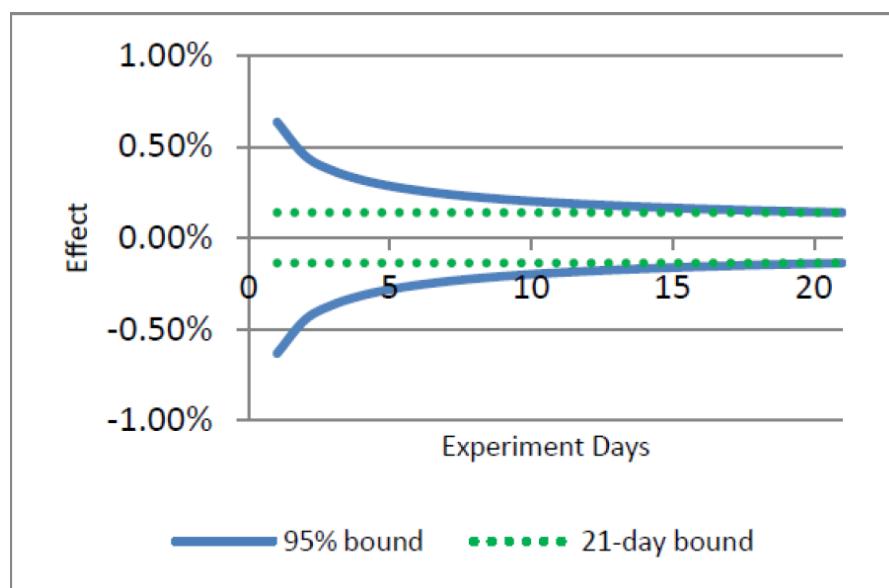
INTERPRETING EXPERIMENT OUTCOMES (1)

- A change was not tested initially. But based on the trend, are we likely to see a positive effect after the initial period?
- If the insignificant effect initially is due to primacy effects, then after the primacy effects wear off and users accept the new feature, the effect can become positive



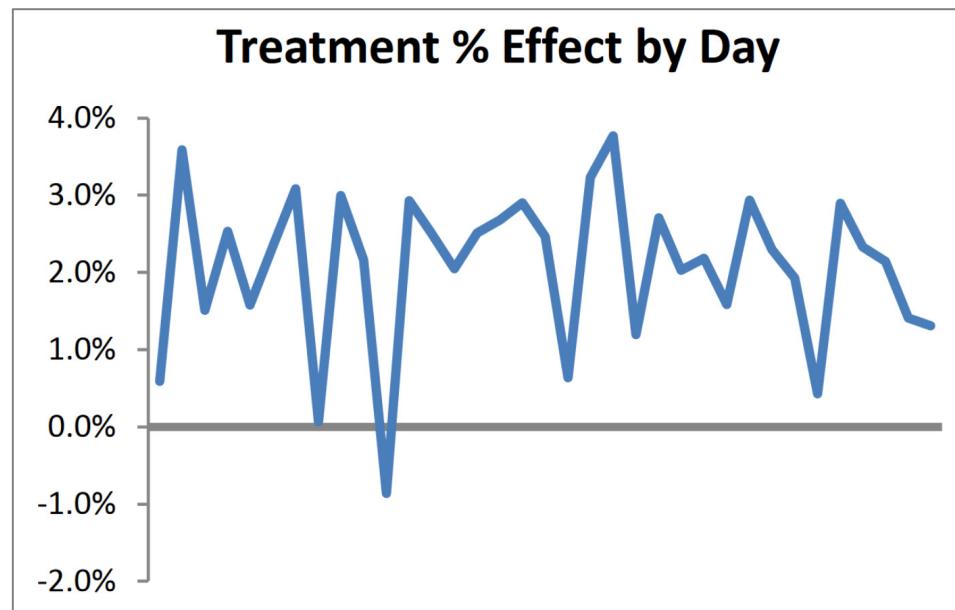
INTERPRETING EXPERIMENT OUTCOMES (1)

- The sample size is small initially, so the initial variance of the cumulative effect is high, i.e., the first few days are highly variable and therefore the effect in the initial days can be much higher or lower than the stabilized effect
- As the sample size increases, the variance reduces, and the cumulative effect gradually converges to the mean, so it appears to trend.



INTERPRETING EXPERIMENT OUTCOMES (2)

- One experiment showed treatment was 4% worse than control. Here is plot of treatment percentage effect by day:



What's going on?

INTERPRETING EXPERIMENT OUTCOMES (2)

- Simpson's paradox (a phenomenon in probability and statistics): a trend appears in several different groups of data but disappears or reverses when these groups are combined

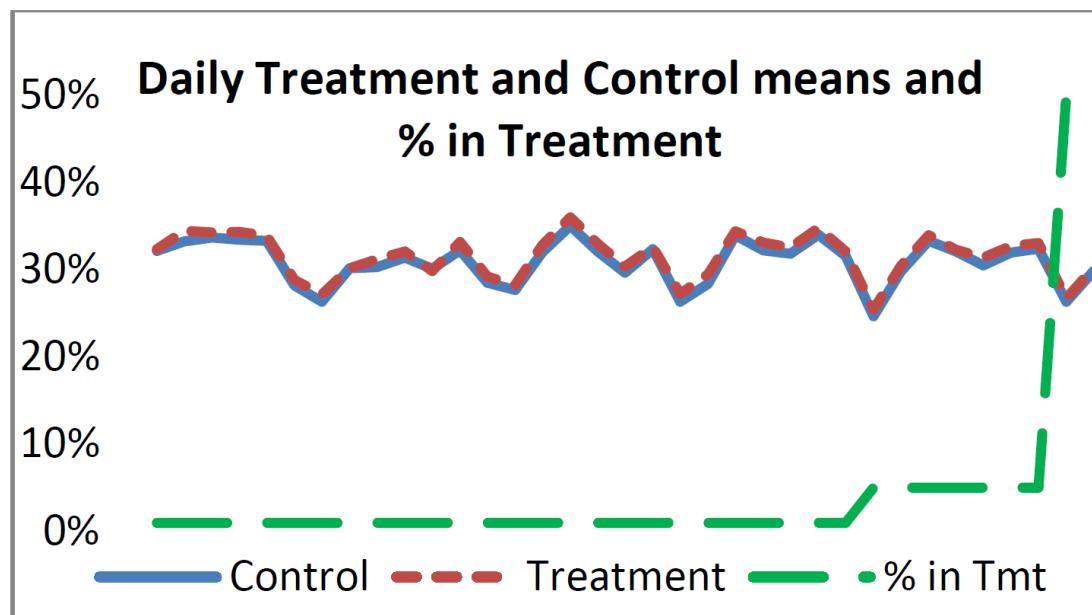


Figure 3 Treatment and Control means and ramp-up percentage for Treatment

INTERPRETING EXPERIMENT OUTCOMES (2)

- Simpson's paradox

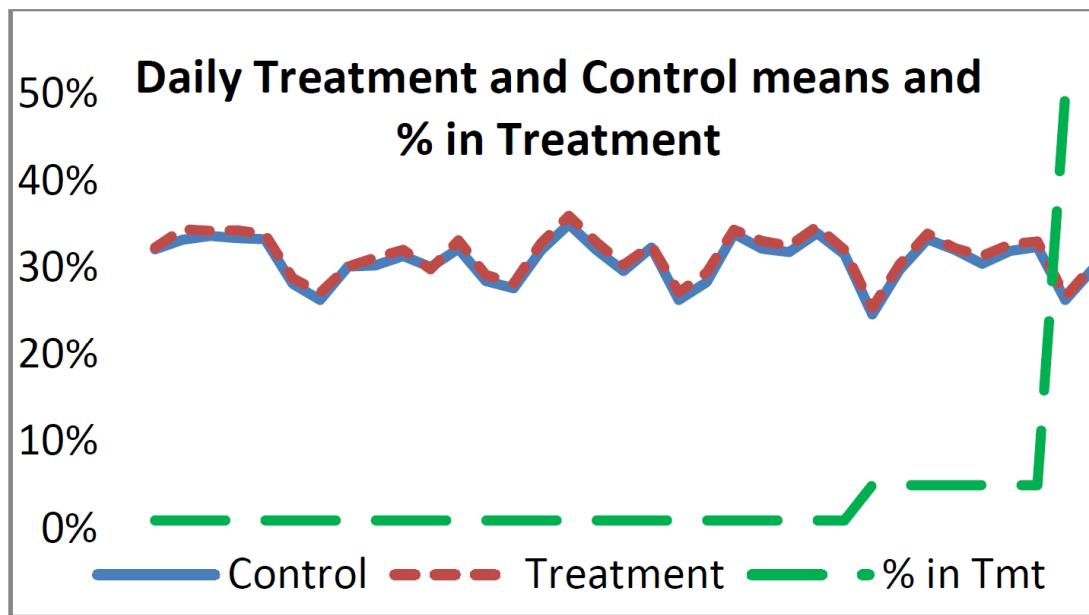


Figure 3 Treatment and Control means and ramp-up percentage for Treatment

To fix it: ensure that relative percentage of users in test and control do not change over the duration of the experiment. Hence, ramp up period must be completed prior to collection of experimental data

INTERPRETING EXPERIMENT OUTCOMES (3)



- Designing a new homepage (bottom one) for office online. Idea was to induce more users to click-through to the download center where they either purchase or download a trial version of office
- No. of Clicks went down by 64%
- Why?

INTERPRETING EXPERIMENT OUTCOMES (3)



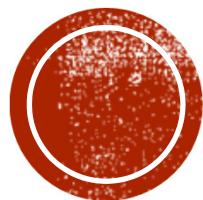
- The big “buy now” button in treatment came with price information
- In control, the call was “Try 2007 for free” and “Buy now”
- Offer of free has huge psychological advantage
- Lack of price information meant that many users in control clicked just to get price information
- We need to pay attention to the final outcome!

MORE LESSONS LEARNED

- Make frequent use of A/A experiment to check consistency
- Segment the results by key attributes such as browser and conduct data quality checks
- Early results are often misleading; repeat the experiments multiple times
- Better understanding of user behavior and software and experimental design

“Getting numbers is easy, getting numbers you can trust is quite difficult.”

-- Ron Kohavi and Roger Longbotham



IV. ATTRIBUTION



ATTRIBUTION EXAMPLE (1)

The screenshot shows the CloudSponge homepage. At the top, there's a navigation bar with links for Home, Product, Price, Demo, and Portfolio. A 'Sign In' button is located in the top right corner. The main heading is 'Discover the Power of a Viral Loop' with the subtext 'Enabling viral growth for social communities'. Below this, there are three sections: 'Import Contacts' (with icons for Yahoo!, Windows Live, Gmail, AOL, and plaxo), 'Send Invites' (with an envelope icon), and 'Go Viral' (with a bar chart and an upward arrow). A large orange 'Sign Up' button is positioned at the bottom of the main content area. At the very bottom, there's a copyright notice and links to Blog, Developer, Contact, Terms, Privacy, and FAQ.

This screenshot shows the same CloudSponge homepage as above, but with several changes. The main heading has been updated to 'Your users have friends. We give you address books.' The 'Import Contacts' section now includes a 'nexopia' logo and a 'DISCOVER' badge. The 'Why CloudSponge' and 'Why Address Books' sections have been expanded with more bullet points. The footer has been simplified, and social media icons for Twitter and Facebook are now present in the bottom right corner.

ATTRIBUTION EXAMPLE (1)

The screenshot shows the CloudSponge homepage with a clean, minimalist design. At the top, there's a navigation bar with links for Home, Product, Price, Demo, and Portfolio. On the right side of the header is a 'Sign In' button. The main content area features a large title 'Discover the Power of a Viral Loop' with a subtitle 'Enabling viral growth for social communities'. Below this, there are three sections: 'Import Contacts' (with icons for Yahoo!, Windows Live, Gmail, AOL, and plaxo), 'Send Invites' (with an envelope icon), and 'Go Viral' (with a bar chart and an upward arrow). A prominent orange 'Sign Up' button is located at the bottom of the main content area. The footer contains copyright information and links to Blog, Developer, Contact, Terms, Privacy, and FAQ.

A Clean Look

The screenshot shows the same CloudSponge homepage as above, but with a different visual style. It features a large banner with the text 'Your users have friends. We give you address books.' and an image of a nexopia login screen. Below this, there are two columns: 'Why CloudSponge' and 'Why Address Books', each with a list of benefits. The footer is identical to the first screenshot.

More excitement

ATTRIBUTION EXAMPLE (1)

The screenshot shows the CloudSponge homepage with a clean, minimalist design. At the top, there's a navigation bar with links for Home, Product, Price, Demo, and Portfolio. On the right side of the header is a 'Sign In' button. The main content area features a large title 'Discover the Power of a Viral Loop' with a subtitle 'Enabling viral growth for social communities'. Below this, there are three sections: 'Import Contacts' (with a Yahoo!, Windows Live, Gmail, AOL, and plaxo icon), 'Send Invites' (with an envelope icon), and 'Go Viral' (with a bar chart and an upward arrow). A prominent orange 'Sign Up' button is located at the bottom of the main content area. The footer contains copyright information and links to Blog, Developer, Contact, Terms, Privacy, and FAQ.

A Clean Look

The screenshot shows the same CloudSponge homepage as above, but with a large red starburst graphic overlaid on the right side containing the text '33% higher TCR'. The rest of the page layout is identical to the first screenshot, featuring the 'Discover the Power of a Viral Loop' section, the 'Import Contacts', 'Send Invites', and 'Go Viral' sections, and the orange 'Sign Up' button. The footer is also present at the bottom.

More excitement

ATTRIBUTION EXAMPLE (2)

World Class Hosting

Safe, Secure and Reliable Hosting

- ▶ 6000 MB of Storage Space
- ▶ 60000 MB Of Monthly Transfer
- ▶ CPANEL with FANTASTICO
- ▶ Unlimited Email Addresses
- ▶ PHP5/Perl/Ruby



With our 30-Day **Money Back Guarantee** you have nothing to lose! Have Questions? Call us toll-free at 1-800-859-8803 Today!

1 **2**

World Class Hosting

Safe, Secure and Reliable Hosting

- ▶ 6000 MB of Storage Space
- ▶ 60000 MB Of Monthly Transfer
- ▶ CPANEL with FANTASTICO
- ▶ Unlimited Email Addresses
- ▶ PHP5/Perl/Ruby



With our 30-Day **Money Back Guarantee** you have nothing to lose! Have Questions? Call us toll-free at 1-800-859-8803 Today!

1 **2**

ATTRIBUTION EXAMPLE (2)

World Class Hosting
Safe, Secure and Reliable Hosting

- ▶ 6000 MB of Storage Space
- ▶ 60000 MB Of Monthly Transfer
- ▶ CPANEL with FANTASTICO
- ▶ Unlimited Email Addresses
- ▶ PHP5/Perl/Ruby



With our 30-Day **Money Back Guarantee** you have nothing to lose! Have Questions? Call us toll-free at 1-800-859-8803 Today!

1 2

Global reach image

World Class Hosting
Safe, Secure and Reliable Hosting

- ▶ 6000 MB of Storage Space
- ▶ 60000 MB Of Monthly Transfer
- ▶ CPANEL with FANTASTICO
- ▶ Unlimited Email Addresses
- ▶ PHP5/Perl/Ruby



With our 30-Day **Money Back Guarantee** you have nothing to lose! Have Questions? Call us toll-free at 1-800-859-8803 Today!

1 2

Security image

ATTRIBUTION EXAMPLE (2)

World Class Hosting
Safe, Secure and Reliable Hosting

- ▶ 6000 MB of Storage Space
- ▶ 60000 MB Of Monthly Transfer
- ▶ CPANEL with FANTASTICO
- ▶ Unlimited Email Addresses
- ▶ PHP5/Perl/Ruby

With our 30-Day **Money Back Guarantee** you have nothing to lose! Have Questions? Call us toll-free at 1-800-859-8803 Today!

1 2

Global reach image

World Class Hosting
Safe, Secure and Reliable Hosting

- ▶ 6000 MB of Storage Space
- ▶ 60000 MB Of Monthly Transfer
- ▶ CPANEL with FANTASTICO
- ▶ Unlimited Email Addresses
- ▶ PHP5/Perl/Ruby

With our 30-Day **Money Back Guarantee** you have nothing to lose! Have Questions? Call us toll-free at 1-800-859-8803 Today!

1 2

Security image



ATTRIBUTION EXAMPLE (3)

L'axelle
Wohlfühl-Pads für die Achseln

Put an end to sweat marks!
Underarm sweat pads - the long duration protection against sweat

Everything under control

- Long duration protection against unpleasant odor building, based on absorbing soft tissue
- Guaranteed maximum freedom of agility
- Invisible and less than 1 mm or 1/26th of an inch
- Perfect fit based on unique fold lines
- Doesn't slip, secure hold

L'axelle underarm sweat pads 30 piece pack
30 underarm sweat pads for 15 utilization units
Freshness and care for you and your clothing

Product in stock

Quantity: 30 pieces pack - € 9,49

9,49 €
+ incl. VAT plus shipping

PLACE ITEM IN eCART

L'axelle
Wohlfühl-Pads für die Achseln

Feel fresh without sweat marks
Underarm sweat pads - a quick help against sweat

Simply feel relaxed.

- Maximum freedom of arm movement through ultra thin feather-light pads
- Perfect fit to your body thanks to the unique fold line
- Unusually high wearing comfort - doesn't slip, reliable hold
- Long duration protection without unpleasant odor building, thanks to the absorbent soft tissue
- Protects clothing

L'axelle underarm sweat pads 30 piece pack
30 underarm sweat pads for 15 utilization units
Freshness and care for you and your clothing

Product in stock

Quantity: 30 pieces pack - € 9,49

9,49 €
+ incl. VAT plus shipping

PLACE ITEM IN eCART

ATTRIBUTION EXAMPLE (3)

L'axelle
Wohlfühl-Pads für die Achseln

Put an end to sweat marks!
Underarm sweat pads - the long duration protection against sweat

Everything under control

- Long duration protection against unpleasant odor building, based on absorbing soft tissue
- Guaranteed maximum freedom of agility
- Invisible and less than 1 mm or 1/26th of an inch
- Perfect fit based on unique fold lines
- Doesn't slip, secure hold

L'axelle underarm sweat pads 30 piece pack
30 underarm sweat pads for 15 utilization units
Freshness and care for you and your clothing

Product in stock

Quantity: 30 pieces pack - € 9,49

9,49 €
= incl. VAT plus shipping

PLACE ITEM IN eCART

L'axelle
Wohlfühl-Pads für die Achseln

Feel fresh without sweat marks
Underarm sweat pads - a quick help against sweat

Simply feel relaxed.

- Maximum freedom of arm movement through ultra thin feather-light pads
- Perfect fit to your body thanks to the unique fold line
- Unusually high wearing comfort - doesn't slip, reliable hold
- Long duration protection without unpleasant odor building, thanks to the absorbent soft tissue
- Protects clothing

L'axelle underarm sweat pads 30 piece pack
30 underarm sweat pads for 15 utilization units
Freshness and care for you and your clothing

Product in stock

Quantity: 30 pieces pack - € 9,49

9,49 €
= incl. VAT plus shipping

PLACE ITEM IN eCART

Loss aversion –
preventing a loss

Gain promotion – promoting
something to be gained

ATTRIBUTION EXAMPLE (3)

93% higher TCR

L'axelle
Wohlfühl-Pads für die Achseln

Put an end to sweat marks!

Underarm sweat pads - the long duration protection against sweat

Everything under control

- Long duration protection against unpleasant odor building, based on absorbing soft tissue
- Guaranteed maximum freedom of agility
- Invisible and less than 1 mm or 1/26th of an inch
- Perfect fit based on unique fold lines
- Doesn't slip, secure hold

GARANTIE QUALITÉ TESTÉ PAR LES EXPERTS

L'axelle underarm sweat pads 30 piece pack
30 underarm sweat pads for 15 utilization units
Freshness and care for you and your clothing

Product in stock

9,49 €
= incl. VAT plus shipping

Quantity: 30 pieces pack - € 9,49 €

L'axelle
Wohlfühl-Pads für die Achseln

Feel fresh without sweat marks

Underarm sweat pads - a quick help against sweat

Simply feel relaxed.

- Maximum freedom of arm movement through ultra thin feather-light pads
- Perfect fit to your body thanks to the unique fold line
- Unusually high wearing comfort - doesn't slip, reliable hold
- Long duration protection without unpleasant odor building, thanks to the absorbent soft tissue
- Protects clothing

GARANTIE QUALITÉ TESTÉ PAR LES EXPERTS

L'axelle underarm sweat pads 30 piece pack
30 underarm sweat pads for 15 utilization units
Freshness and care for you and your clothing

Product in stock

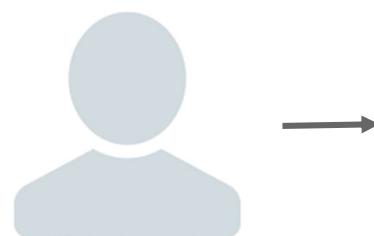
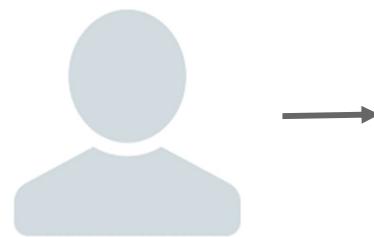
9,49 €
= incl. VAT plus shipping

Quantity: 30 pieces pack - € 9,49 €

Loss aversion –
preventing a loss

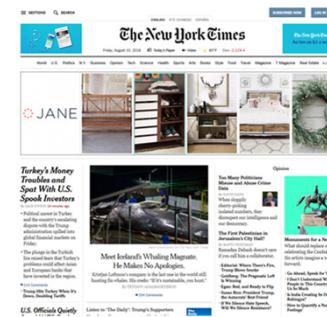
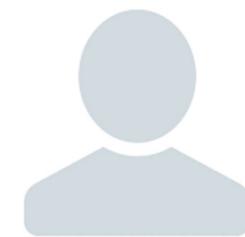
Gain promotion – promoting
something to be gained

ATTRIBUTION EXAMPLE (4)



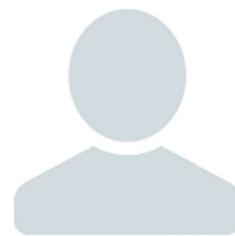
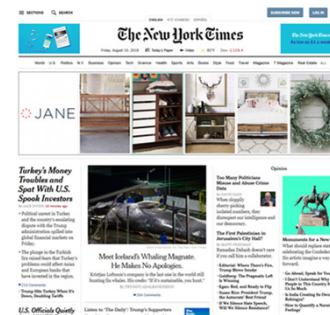
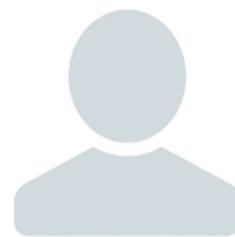
Compare those who saw your advertisement with those who did not see your ad to determine if your ad had a positive effect.

ATTRIBUTION EXAMPLE (4)



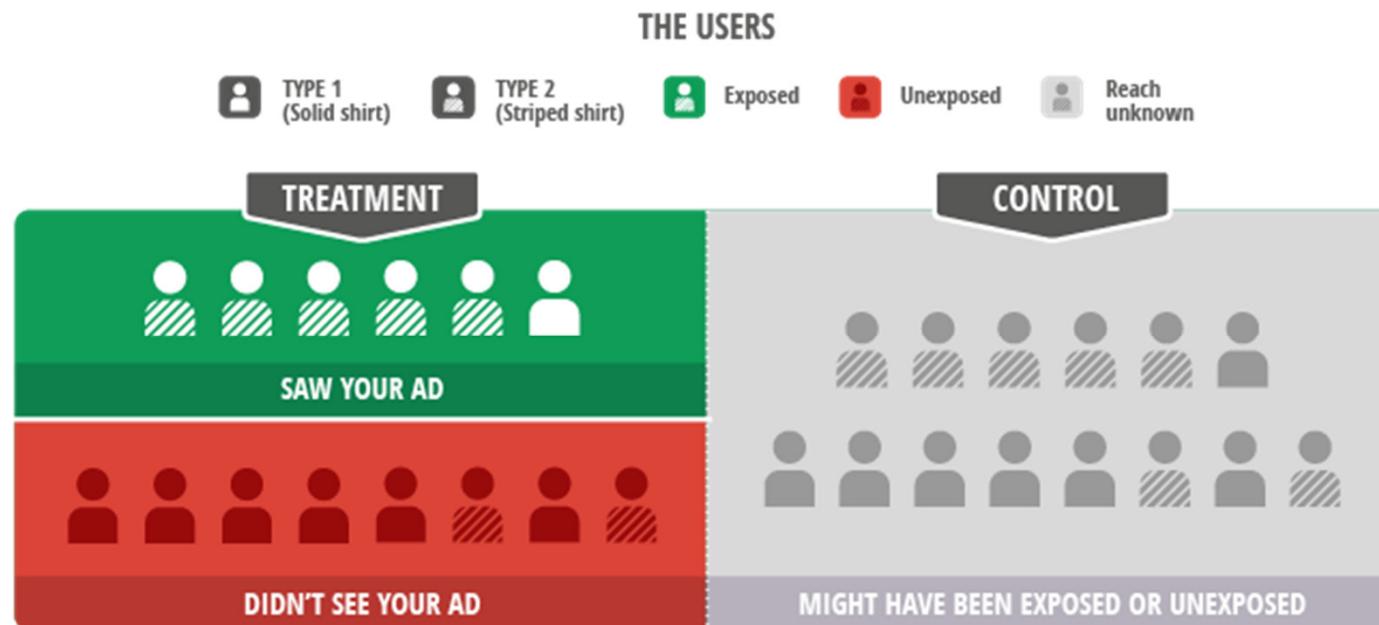
- But whether a user sees a particular ad depends on their behavior, competitive bidding among advertisers, and numerous targeting parameters
- What if the population who saw your ad is different from the population who didn't see your ad?

ATTRIBUTION EXAMPLE (4)



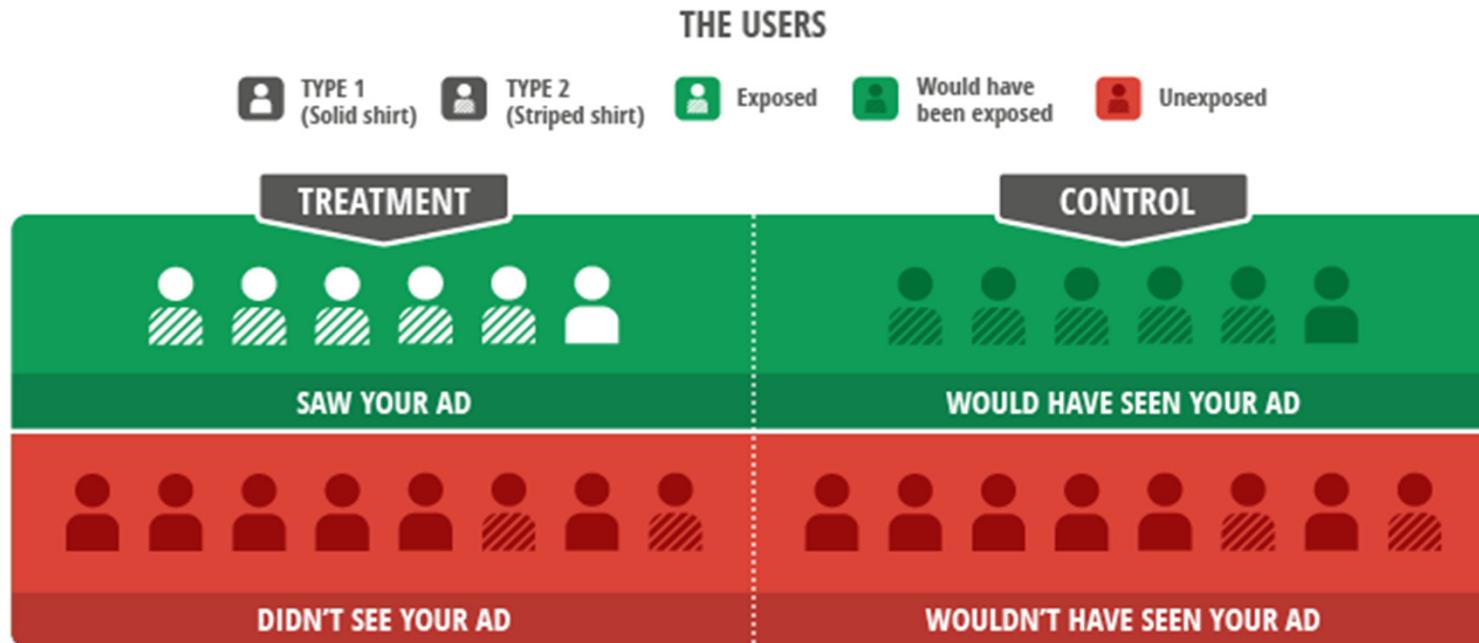
One method of conducting an experiment to measure ad effectiveness is to create two different ads: one for your brand and another for something entirely unrelated to your business. Put both ads in your ad group, so Google will randomly assign people to one ad or the other. Standard practice, with some downsides.

SIMPLY RANDOM



- Simply divide audience into two groups and show ads to only one. Compare behavior of all users, not just those who saw an ad, between the two groups to measure the causal effect of the advertising.
- Fair comparison, but a lot of noise.

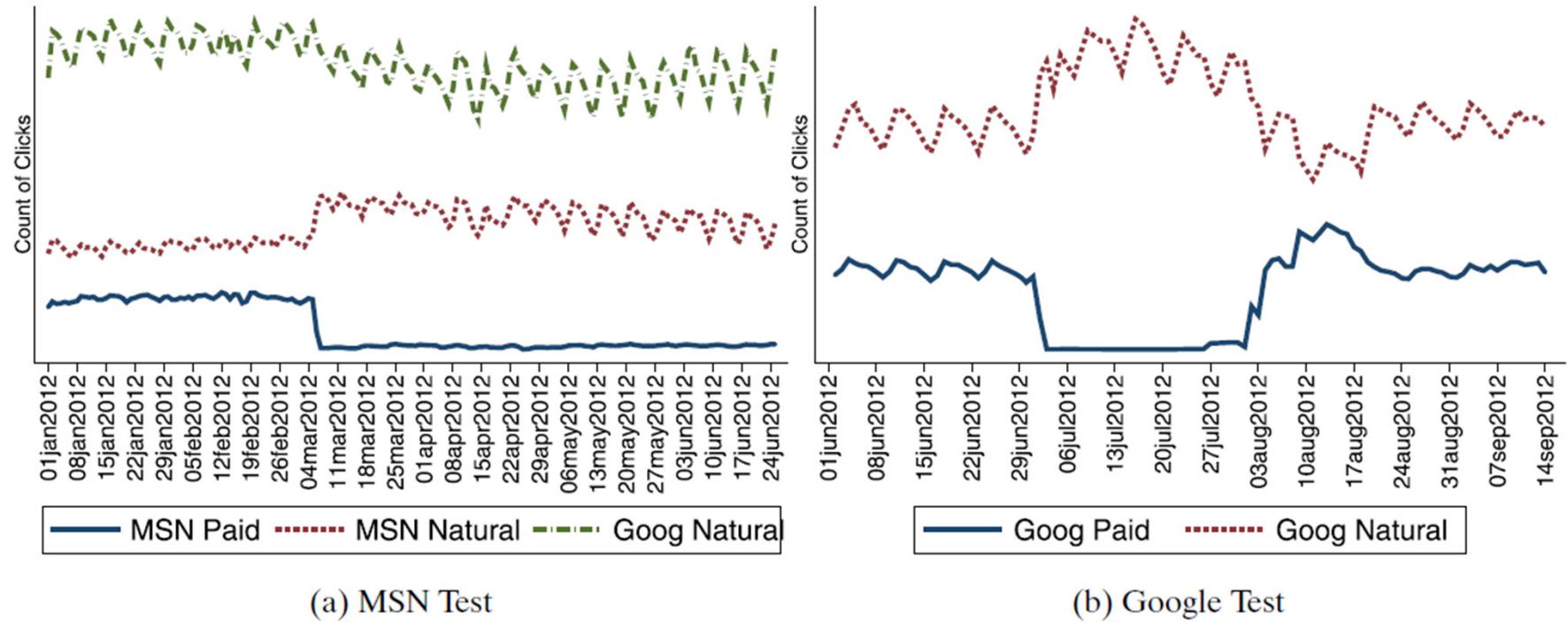
“GHOST ADS”



- Ghost ads methodology lets advertisers run an experiment without spending money on irrelevant ads – every time an advertiser's ad is selected to run (based on its usual auction method), it then flips a coin to decide whether to run the ad or not to run the ad. When the ad doesn't run, the advertiser doesn't have to pay for that ad exposure.
- Less noise.
- It measures against the relevant competitive baseline: would have been exposed users when ads are turned off.

PAID SEARCH EFFECTIVENESS FOR EBAY BRANDED KEYWORDS

PAID SEARCH EFFECTIVENESS



(a) MSN Test

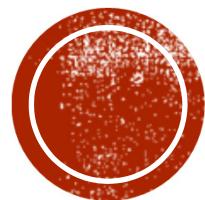
(b) Google Test

Not employing search ads on branded keywords, eBay lost only .529% of traffic, meaning they retained 99.5% of traffic

– Thomas Blake, Chris Nosko, and Steven Tadelis

HETEROGENEOUS TREATMENT EFFECTS

- The average treatment effect (ATE) is the primary metric reported in A/B testing.
- Understanding conditional average treatment effects (CATE), where we condition on some segment, provides insights (e.g., treatment works better in some markets) and can help isolate bugs (e.g., treatment fails for Firefox version X).
- What are good segments to condition?
 - Market/Country
 - User activity level (e.g., heavy, light, new)
 - Device/platform (e.g., desktop, mobile, browser and version, OS and version)
 - Day of week and time
 - Product specific segments (e.g., Netflix segments by network speed; Airbnb segments by first booking and vacation destination)
- Correlations vs. Causation (e.g., different CATE for iOS vs. Android may be due to population, not the implementation)



V. ACT UPON FINDINGS



CAPTURING LOST CONVERSIONS

Non-converters:

- Did not like what they saw
- Were not likely to convert anyway
- Or, were not ready to convert (but may be ready in the future)

CAPTURING LOST CONVERSIONS

Non-converters:

- Did not like what they saw
 - change design
- Were not likely to convert anyway
 - minimize traffic or make ad revenue from their visit
- Or, were not ready to convert (but may be ready in the future)
 - generic searches may naturally produce later branded searches followed by conversions
 - engage in retargeting

GENERIC-TO-DIRECT SPILLOVER EXAMPLE: AN AUTOMOTIVE WEBSITE

- The purchase cycle for an automobile is quite large. Most consumers are unlikely to purchase on their first visit
 - Of 3,186 traffic-producing keywords, 599 produced later increases in direct traffic
 - Company keywords (branded keywords) are the most likely to produce later direct visits.
 - Search keywords (e.g., “find Nissan Altima”), comparison keywords (e.g., “compare midsize cars”), and information keywords are likely to produce later return visits.
 - Mileage, Condition, and Price related keywords produced a lower likelihood of repeat visits, suggesting that consumers who were closer to the purchase decision were less likely to make a return visit.
- Oliver J. Rutz, Michael Trusov, and Randolph E. Bucklin

DYNAMIC RETARGETING

View specific product on website

The screenshot shows the Sarenza.co.uk homepage with a navigation bar for WOMEN, MEN, KIDS, BAGS, and BRANDS. A banner at the top states "Deliveries are ALWAYS free... even during the Sale". Below this, a link to "Kids' Sale" and "Geox Sale" is shown. The main focus is a large image of a pink Geox children's shoe with cartoon cat graphics. To the right of the shoe, there is a "20%" discount badge, a five-star rating, and options to "Select colour" (pink) and "Select size" (12.5K or 13.5K). A "Size chart" link is also present.

Retargeted by product-specific ad

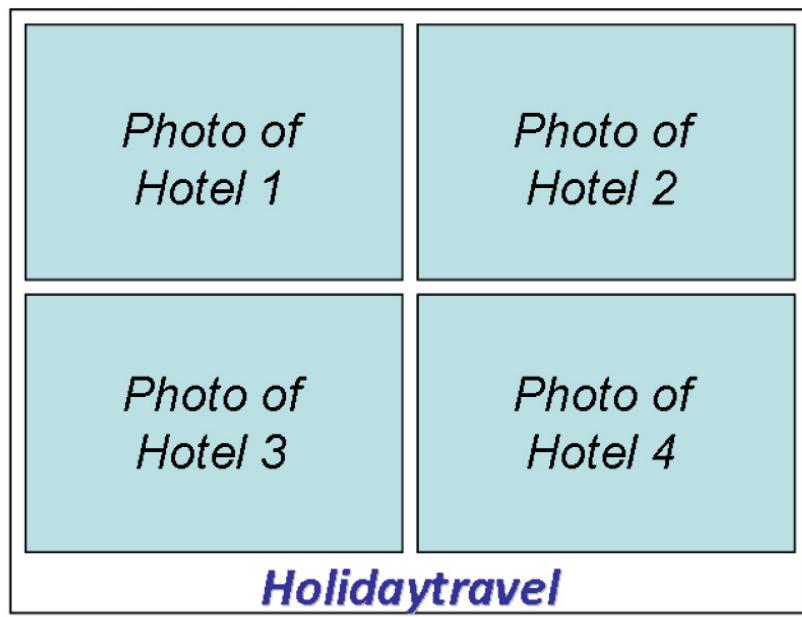
The screenshot shows a retargeted advertisement on the Sarenza.co.uk website. The ad features a grid of four pink Geox children's shoes. The top-right shoe, which is identical to the one shown in the first screenshot, is circled in red. The ad includes a "Free delivery on all orders + Free 100 days returns policy" banner at the top. Navigation arrows are visible on the left and right sides of the grid.

AN DYNAMIC RETARGETING EXPERIMENT

Generic Retargeting

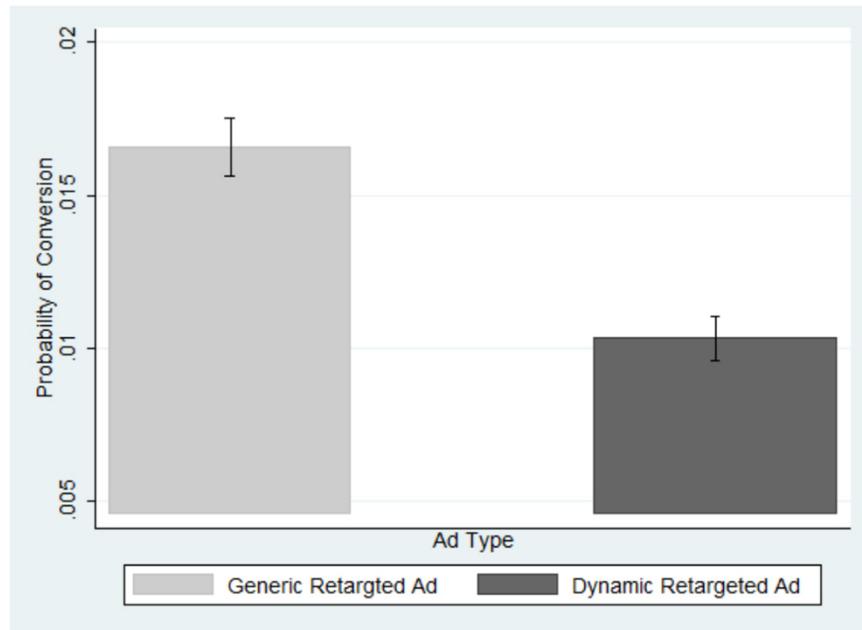


Dynamic Retargeting
(Specific)

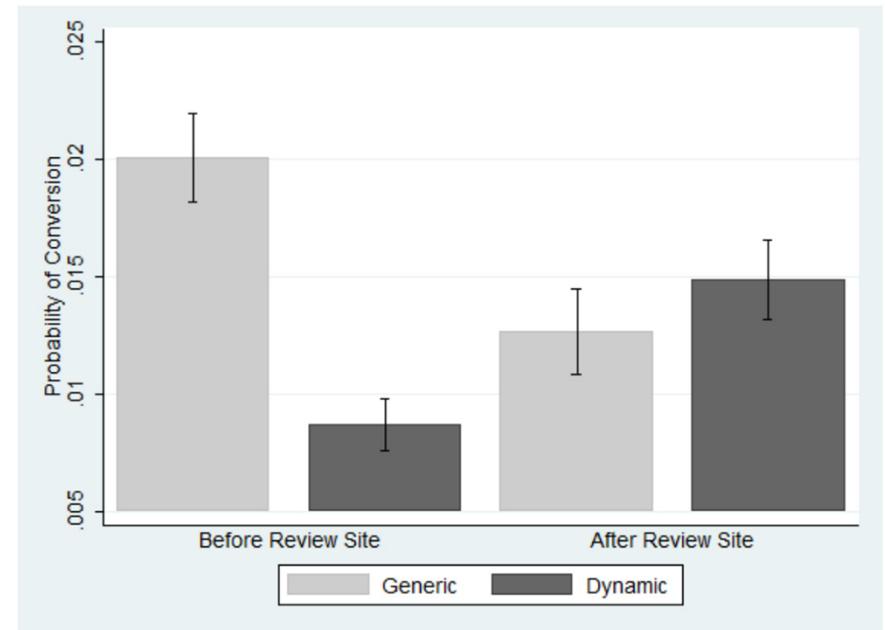


ads show pictures of the specific hotel investigated by the consumer.

EXPERIMENT RESULTS: PHASE OF THE CUSTOMER JOURNAL MATTERS



Comparison of Conversion for Generic vs Dynamic Ad Exposure



Comparison of Conversion for Generic vs Dynamic Ad Exposure: Sample Restricted to Those Who Visited A Review Site

– Anja Lambrecht and Catherine Tucker

BROAD VS. NARROW CONSTRUAL

