

A/B Testing

Crowdsourcing and Human Computation

Lecture 23

Instructor: Chris Callison-Burch

TA: Ellie Pavlick

Website: crowdsourcing-class.org

Active versus Passive Crowdsourcing

- So far we have mainly looked at *active* crowdsourcing, where we explicitly solicit help from the crowd
- Many applications of crowdsourcing rely on *passive* information collection from multitudes of individual

Example: Apple Maps

- iOS allows users to “help improve maps” by enabling a feature called “frequent locations”
- Frequent locations gives Apple a method to verify business locations and other destinations by tracking user movements in the aggregate
- Participation also transmits drive and other travel time data to Apple

A/B Testing

- A/B Split Testing is a mechanism for passive crowdsourcing that allows web developers to empirically optimize the design of their sites
- Splits web users into two groups and shows them slightly different versions of the site
- Measures the behavior of the groups in aggregate and calculates whether one design leads to a better measurable outcome

Why A/B tests?

- Lets us evaluate the goodness of alternate designs, instead of relying on our intuitions
- A typical web site may convert only 2% of its visitors into customers
- Small changes can have a big impact
- Google uses A/B testing all the time, and makes it available through Google Analytics

What sorts of things can you optimize with A/B tests?

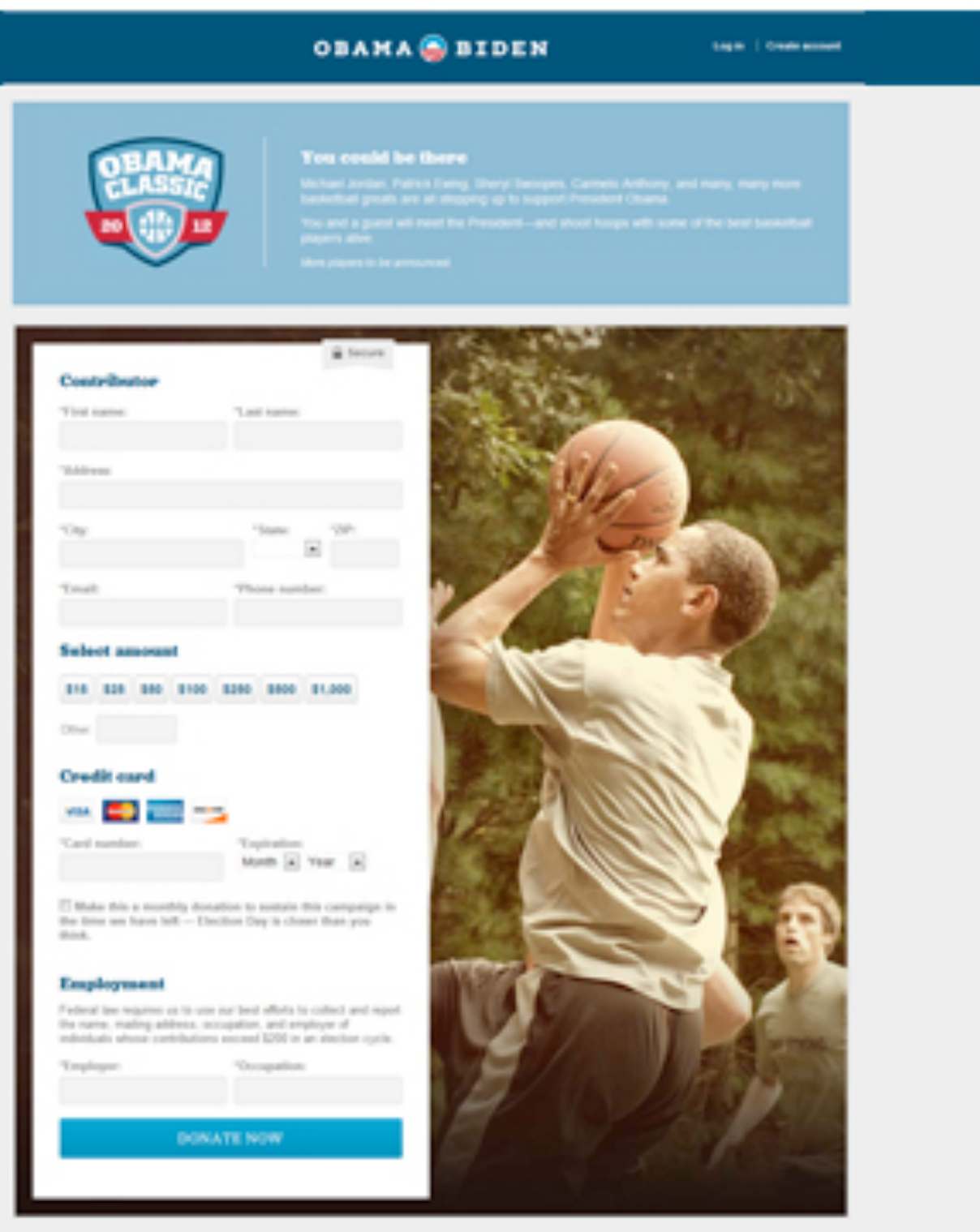
- Whether changing the order of collecting form information gets users to stick through to the end
- Whether changing the copywriting on your page improves things
- Whether different images are better at motivating web site visitors to do something that you want them to

What outcomes could
you measure?

A/B testing was used to optimize the Obama Campaign

- Kyle Rush was the deputy director of frontend web development at Obama for America
- Managed online fundraising totaling \$690 million in 20 months
- Conducted 500+ A/B tests, which increased the donation conversion rate by 49% and the email acquisition conversion rate by 161%

CONTROL



“SEQUENTIAL”



↑ +5%

<http://kyberusak>

CONTROL

OBAMA BIDEN

DINNER WITH BARACK

Your chance to meet the President



No purchase, payment, or contribution necessary to enter or win. Contributing will not improve chances of winning. Void where prohibited. Entries must be received by September 26, 2012. You may enter by contributing to Obama Victory Fund 2012 here or click here to enter without contributing. Three winners will each receive the following prize package: round-trip tickets for winner from within the 48 U.S. States, DC, or Puerto Rico to a destination to be determined by the Sponsor; hotel accommodations; and dinner with President Obama on a date to be determined by the Sponsor (approximate retail value of all prizes: \$4,800). Odds of winning depend on number of entries received. Promotion open only to U.S. citizens, or lawful permanent U.S. residents who are legal residents of 48 United States, District of Columbia and Puerto Rico, and 18 or older (or age of majority under applicable law). Promotion subject to Official Rules. Official rules and additional restrictions on eligibility. Sponsor: Obama for America, 1300 S. Randolph St., Chicago, IL 60607.

OBAMA BIDEN

[Privacy Policy](#) [Terms of Service](#)

Contributions or gifts to Obama Victory Fund 2012 are not tax deductible.

PAID FOR BY OBAMA VICTORY FUND 2012, A JOINT FUNDRAISING COMMITTEE AUTHORIZED BY OBAMA FOR AMERICA, THE DEMOCRATIC NATIONAL COMMITTEE, AND THE STATE DEMOCRATIC PARTIES IN THE FOLLOWING STATES: CO, FL, IL, IN, MI, NH, NC, OH, PA, VA, AND WA.

© 2011-2012 Obama for America. All Rights Reserved.

IMAGE VARIATION

OBAMA BIDEN

DINNER WITH BARACK

Your chance to meet the President



No purchase, payment, or contribution necessary to enter or win. Contributing will not improve chances of winning. Void where prohibited. Entries must be received by September 26, 2012. You may enter by contributing to Obama Victory Fund 2012 here or click here to enter without contributing. Three winners will each receive the following prize package: round-trip tickets for winner from within the 48 U.S. States, DC, or Puerto Rico to a destination to be determined by the Sponsor; hotel accommodations; and dinner with President Obama on a date to be determined by the Sponsor (approximate retail value of all prizes: \$4,800). Odds of winning depend on number of entries received. Promotion open only to U.S. citizens, or lawful permanent U.S. residents who are legal residents of 48 United States, District of Columbia and Puerto Rico, and 18 or older (or age of majority under applicable law). Promotion subject to Official Rules. Official rules and additional restrictions on eligibility. Sponsor: Obama for America, 1300 S. Randolph St., Chicago, IL 60607.

OBAMA BIDEN

[Privacy Policy](#) [Terms of Service](#)

Contributions or gifts to Obama Victory Fund 2012 are not tax deductible.

PAID FOR BY OBAMA VICTORY FUND 2012, A JOINT FUNDRAISING COMMITTEE AUTHORIZED BY OBAMA FOR AMERICA, THE DEMOCRATIC NATIONAL COMMITTEE, AND THE STATE DEMOCRATIC PARTIES IN THE FOLLOWING STATES: CO, FL, IL, IN, MI, NH, NC, OH, PA, VA, AND WA.

© 2011-2012 Obama for America. All Rights Reserved.

↑ +19%

<http://ky.lerush.es.t>

CONTROL

OBAMA BIDEN

✔ Your donation was successful. Thank you.

Secure

Save your payment information for next time.

Save the payment information from this donation into your existing BarackObama.com account so that you don't have to enter it more than once.

You can manage your saved payment information by clicking your name in the top right when you are logged in.

Email

Enter your BarackObama.com account password

[Forgot your password?](#)

☐ Link my mobile phone number with my account so that I can donate and receive campaign updates via text message. Message and data rates may apply. My mobile number is: [Change this number](#)

☐ Keep me logged in. (If this is a public computer or you are concerned about unauthorized access to this device, you should not enable this setting.)

SAVE PAYMENT INFORMATION

Made just now

OBAMA BIDEN

[Privacy Policy](#) [Terms of Service](#)

PAID FOR BY OBAMA FOR AMERICA

© 2011-2012 Obama for America. All Rights Reserved

SEGUE COPY

OBAMA BIDEN

✔ Your donation was successful. Thank you.

Secure

Now, save your payment information.

Save the payment information from this donation into your existing BarackObama.com account so that you don't have to enter it more than once.

You can manage your saved payment information by clicking your name in the top right when you are logged in.

Email

Enter your BarackObama.com account password

[Forgot your password?](#)

☐ Link my mobile phone number with my account so that I can donate and receive campaign updates via text message. Message and data rates may apply. My mobile number is: [Change this number](#)

☐ Keep me logged in. (If this is a public computer or you are concerned about unauthorized access to this device, you should not enable this setting.)

SAVE PAYMENT INFORMATION

Made just now

OBAMA BIDEN

[Privacy Policy](#) [Terms of Service](#)

PAID FOR BY OBAMA FOR AMERICA

© 2011-2012 Obama for America. All Rights Reserved

↑ +21%

Statistics behind A/B testing

- One commonly used method is the G-test
- Invented by Karl Pearson in 1900
- Provides a method for comparing 2 data sets, or answering a yes/no question
- Good tutorial by Ben Tilly on “Effective A/B Testing”

What to measure?

- Start your A/B test
- Divide your users into groups A and B
- Decide whether each person did what you wanted
- Reduce your results into 4 numbers:
\$a_yes, \$a_no, \$b_yes, \$b_no

Counts of outcomes

	Yes	No	
A	\$a_yes	\$a_no	\$a
B	\$b_yes	\$b_no	\$b
	\$yes	\$no	\$total

Totals

	Yes	No	
A	\$a_yes	\$a_no	\$a
B	\$b_yes	\$b_no	\$b
	\$yes	\$no	\$total

Expectations

	Yes	No	
A	$\$e_a_yes$	$\$e_a_no$	$\$a$
B	$\$e_b_yes$	$\$e_b_no$	$\$b$
	$\$yes$	$\$no$	$\$total$

$$\$e_a_yes = \$a * \$yes / \$total$$

$$\$e_a_no = \$a * \$no / \$total$$

$$\$e_b_yes = \$b * \$yes / \$total$$

$$\$e_b_no = \$b * \$no / \$total$$

G-test

observations

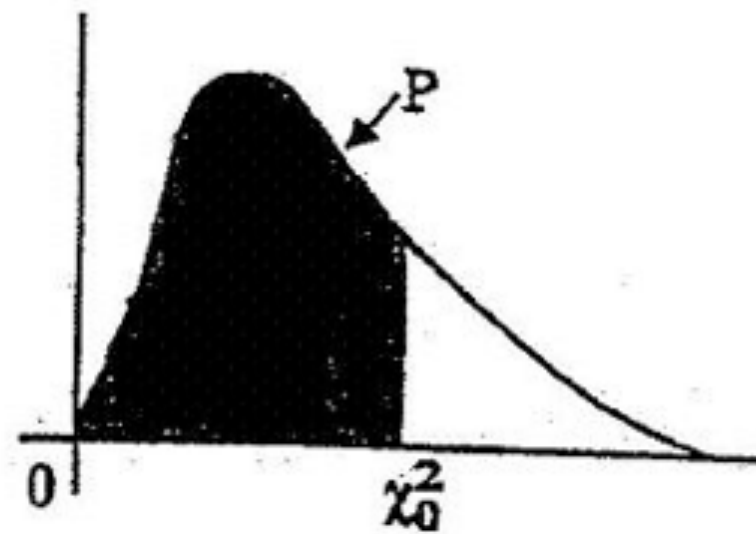
	Yes	No
A	$\$a_yes$	$\$a_no$
B	$\$b_yes$	$\$b_no$

expectations

Yes	No
$\$e_a_yes$	$\$e_a_no$
$\$e_b_yes$	$\$e_b_no$

$$\begin{aligned} \$G_test = & 2 * (\$a_yes * \ln(\$a_yes / \$e_a_yes) \\ & + \$a_no * \ln(\$a_no / \$e_a_no) \\ & + \$b_yes * \ln(\$b_yes / \$e_b_yes) \\ & + \$b_no * \ln(\$b_no / \$e_b_no) \end{aligned}$$

χ^2 Distribution



The table below gives the value x_0^2 for which $P[\chi^2 < x_0^2] = P$ for a given number of degrees of freedom and a given value of P .

Degrees of Freedom	Values of P									
	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	
2	0.01	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	

G-test Calculator

How many versions?

Success is measured by

- ☒ count of successes
☐ percentage of trials that succeeded

Use the Yates' continuity correction? ☒

Test	Trials	Success
<input type="text" value="A"/>	<input type="text" value="1000"/>	<input type="text" value="100"/>
<input type="text" value="B"/>	<input type="text" value="500"/>	<input type="text" value="25"/>

The G-test statistic is 11.0928 so version 'A' wins with 99.91% confidence.

<http://elem.com/~btilly/effective-ab-testing/g-test-calculator.html>

Optimizely

- <http://www.optimizely.com>
- 4 minute video

A/B Split Testing Protocol

- Identify your initial control web page – this could be your current landing page or whatever you want to optimize
- Establish your goals – what is the thing that you want to optimize? Number of people signing up for your service? Revenue generated by a particular ad campaign?

A/B Split Testing Protocol

- Determine how long you need to run the experiment – this depends on how much traffic your web site gets, and what level of statistical significance you want
- Create 1 to 3 significant re-designs – your designers can propose a bunch of different overhauls, use the initial phase to hone in on the best high-level re-design

A/B Split Testing Protocol

- Use A/B testing to choose among the different re-designs. Ideally you can test every pages against every other one, but if that is impractical, you can do a tournament
- Based on the results, choose your true control page – this initial pick will likely generate the lion's share of the improvements

A/B Split Testing Protocol

- Finally, optimize the nitty-gritty elements of the web page using A/B testing
 - Headline
 - Call to Action
 - Page Copy
 - Graphics
 - Color
 - Configuration of Page Elements
 - Etc.

A/B Split Testing Protocol

- Finally, optimize the nitty-gritty elements of the web page using A/B testing
 - Headline
 - Call to Action
 - Page Copy
 - Graphics
 - Color
 - Configuration of Page Elements
 - Etc.

You are part of an experiment

- Who uses A/B testing?
- Pretty much every web site out there
- Google, Amazon, Facebook
- At what point does it become creepy?

Not Creepy

Creepy

Layout of a web site

Manipulating our Facebook feeds
to modify our emotions

Font choice

Dating matches who would be bad
for our tastes

What ads we see (mostly)

Ads for arrest record that are more
strongly associated with African
American names

Companies trying to make a
good product

People at companies playing
social scientists w/o normal
safeguards

Detecting Emotional Contagion in Massive Social Networks

Lorenzo Coviello¹, Yunkyu Sohn², Adam D. I. Kramer³, Cameron Marlow³, Massimo Franceschetti¹, Nicholas A. Christakis^{4,5}, James H. Fowler^{2,6*}

Abstract

Happiness and other emotions spread between people in direct contact, but it is unclear whether massive online social networks also contribute to this spread. Here, we elaborate a novel method for measuring the contagion of emotional expression. With data from millions of Facebook users, we show that rainfall directly influences the emotional content of their status messages, and it also affects the status messages of friends in other cities who are not experiencing rainfall. For every one person affected directly, rainfall alters the emotional expression of about one to two other people, suggesting that online social networks may magnify the intensity of global emotional synchrony.

Introduction

Happiness and other emotions have recently been an important focus of attention in a wide range of disciplines, including psychology, economics, and neuroscience [1,2,3,4]. Some of this work suggests that emotional states can be transferred directly from one individual to another via mimicry and the copying of emotionally-relevant bodily actions like facial expressions [5]. Experiments have demonstrated that people can “catch” emotional states they observe in others over time frames ranging from seconds to months [6,7], and the possibility of emotional contagion between strangers, even those in ephemeral contact, has been documented by the effects of “service with a smile” on customer satisfaction and tipping [8].

Longitudinal data from face-to-face social networks has established that emotions as diverse as happiness [9], loneliness [10], and depression [11] are correlated between socially-connected individuals, and related work suggests that these correlations also exist online [4,12,13,14,15]. However, it is difficult to ascertain whether correlations in observational studies

large-scale experiment, they may have other goals that constrain its design. For example, they may wish to provide a uniform online experience to all users, which reduces their willingness to create experimental treatment groups of sufficient size to take advantage of their massive scale.

Here, we propose an alternative method for detecting emotional contagion in massive social networks that is based on instrumental variables regression, a technique pioneered in economics [23]. In an experiment we would directly control each user’s emotional expression to see what impact it has on their friends’ emotional expression. However, since this is infeasible in our massive-scale setting, we identify a source of variation that directly affects the users’ emotional expression (this variable is called an “instrument”). For this instrument, we use rainfall. Importantly, rainfall is unlikely to be causally affected by human emotional states, so if we find a relationship it suggests that rainfall influences emotional expression and not vice versa. We then measure whether or not the changes induced by the instrument predict changes in the friends’ emotional expression. Instead of changing the user’s

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^b, and Jeffrey T. Hancock^{c,d}

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; ^bCenter for Tobacco Control Research and Education, University of California, San Francisco, CA 94143; and Departments of ^cCommunication and ^dInformation Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social

network, collected over a 20
moods (e.g., depression, ha
networks [Fowler JH, Chris
though the results are contr
who use Facebook, we test
outside of in-person interac
the amount of emotional co
expressions were reduced, i
and more negative posts;
duced, the opposite pattern
emotions expressed by oth
emotions, constituting exp
contagion via social netwo

contrast to prevailing assumptions, in-person interaction and non-verbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.

computer-mediated communication | social media | big data

Emotional states can be transferred to others via emotional contagion, leading them to experience the same emotions as those around them. Emotional contagion is well established in

demonstrated that (i) emotional contagion occurs via text-based computer-mediated communication (7); (ii) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (iii) people's emotional expressions on Facebook predict friends'

though some shared late, however, thereoods are contagiousriencer and target.motions, which areews Feed" productduce much moreFeed filters posts,News Feed is thethat friends share.News Feed is deacebook continuallyviewers the content

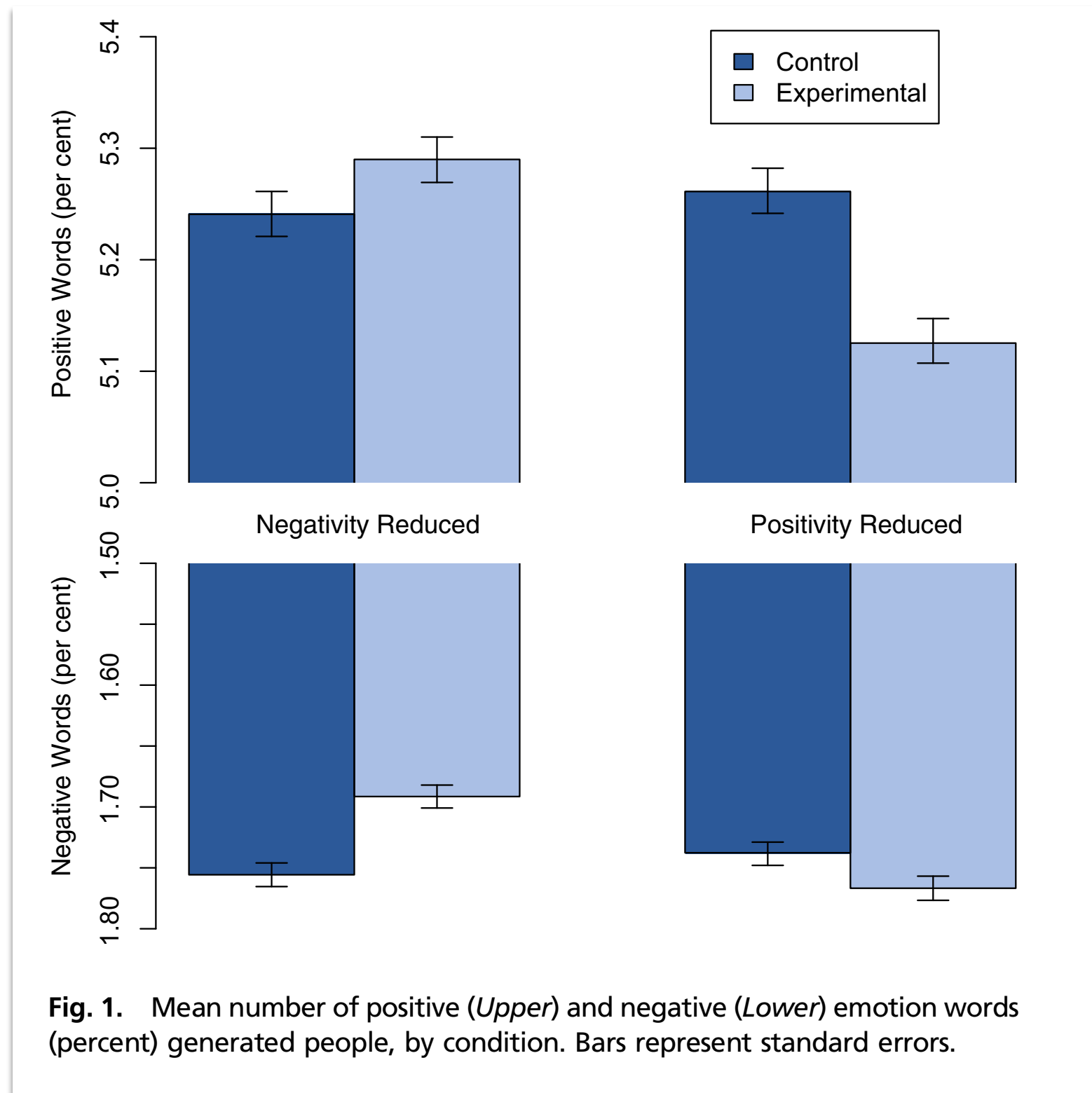
they will find most relevant and engaging. One such test is reported in this study: A test of whether posts with emotional content are more engaging.

The experiment manipulated the extent to which people ($N = 689,003$) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led people to change their own posting behaviors, in particular whether exposure to emotional content led people to post content that was consistent with the exposure—thereby testing whether exposure to verbal affective expressions leads to similar verbal expressions.

Significance

We show, via a massive ($N = 689,003$) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.

Two parallel experiments were conducted for positive and negative emotion: One in which exposure to friends positive emotional content in their News Feed was reduced, and one in which exposure to negative emotional content in their News Feed was reduced.



Senator asks FTC to investigate Facebook's mood study

After the social network altered the news feeds of nearly 700,000 users without telling them, Sen. Mark R. Warner wants to know if there should be oversight on these types of experiments.

by **Dara Kerr**  @darakerr / July 9, 2014 5:32 PM PDT

 4 /  66 /  214 /  11 /  /  more +

The US government might now weigh in on Facebook's secret 2012 study on the moods of nearly 700,000 users.

Sen. Mark R. Warner (D-Va.) has penned a letter ([PDF](#)) to the Federal Trade Commission asking the regulatory agency to investigate the issue. He requested that the FTC look into the ramifications of the experiment and to consider whether rules should be put in place for such types of experiments on social networks.

"I understand that social-media companies are looking for ways to extract value from the information willingly provided by their huge customer base," Warner said in a **statement**. "But I think many consumers were surprised to learn they had given permission by agreeing to Facebook's terms of service. And I think the industry could benefit from a conversation about what are the appropriate rules of the road going forward."

Basic Ethical Principles

1. Respect for Persons – individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled to protection
2. Beneficence – do not harm and maximize possible benefits and minimize possible harms
3. Justice – Who ought to receive the benefits of research and bear its burdens?



Dating Research from OkCupid

We Experiment On Human Beings!

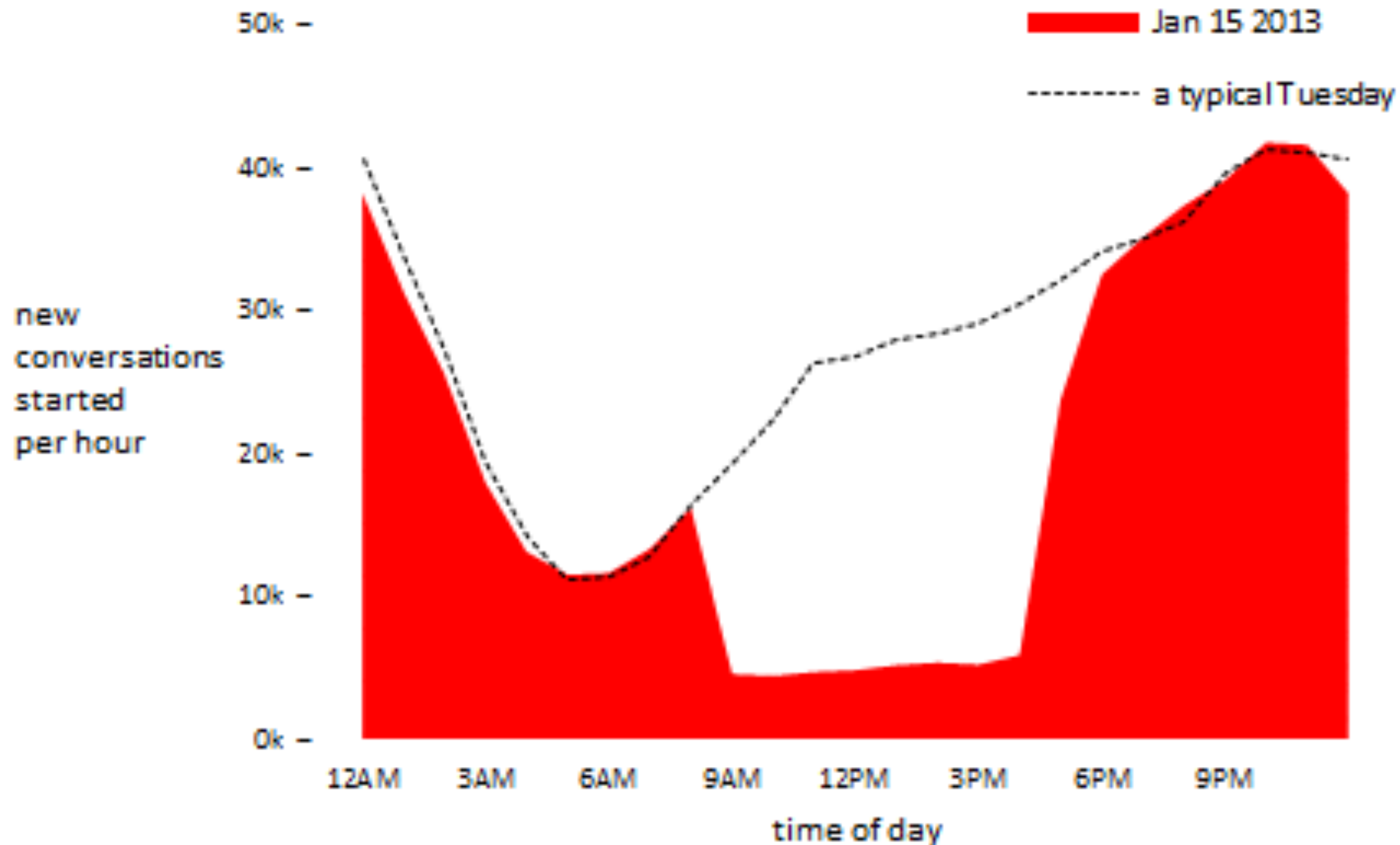
July 28th, 2014 by [Christian Rudder](#)

Tweet 2,831

Share 10k

I'm the first to admit it: we might be popular, we might create a lot of great relationships, we might blah blah blah. But OkCupid doesn't really know what it's doing. Neither does any other website. It's not like people have been building these things for very long, or you can go look up a blueprint or something. Most ideas are bad. Even good ideas could be better. Experiments are how you sort all this out. Like this

Love should be blind

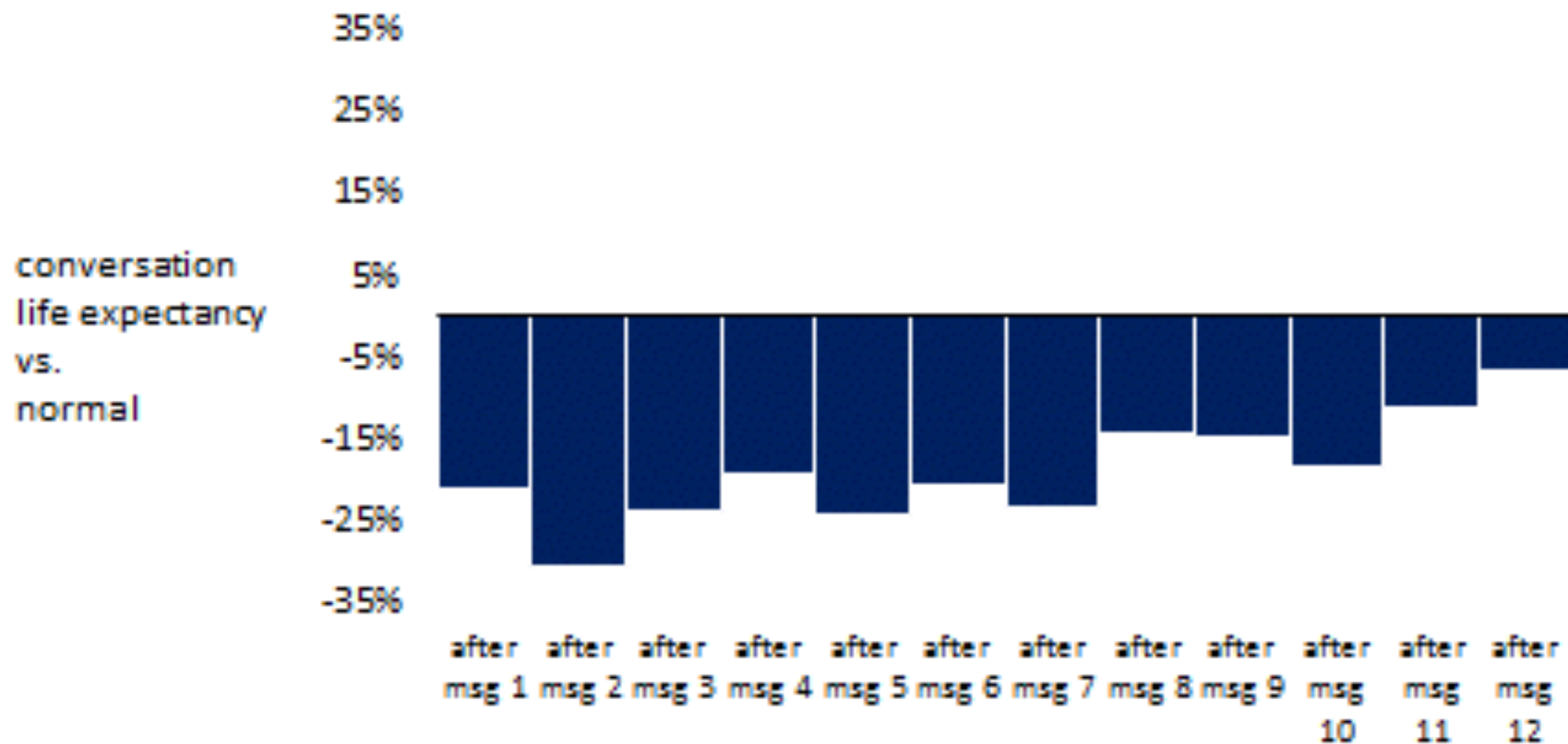


Love should be blind

- people responded to first messages 44% more often
- conversations went deeper
- contact details (phone numbers, emails) were exchanged more quickly
- in short, OkCupid worked better

Love should be blind

What happened to conversations in progress at the end of Love Is Blind Day



where in the thread the photos came back on

Picture is worth 1000 words?

The screenshot shows the OkCupid website interface. At the top, the OkCupid logo is on the left, and language options (Português, Français, English) and links (About Us, Contact Us, Homescreen, Mailbox, Edit Profile, Log Out) are on the right. A red arrow points from the text 'Picture is worth 1000 words?' to the 'look up user' search bar in the top navigation bar.

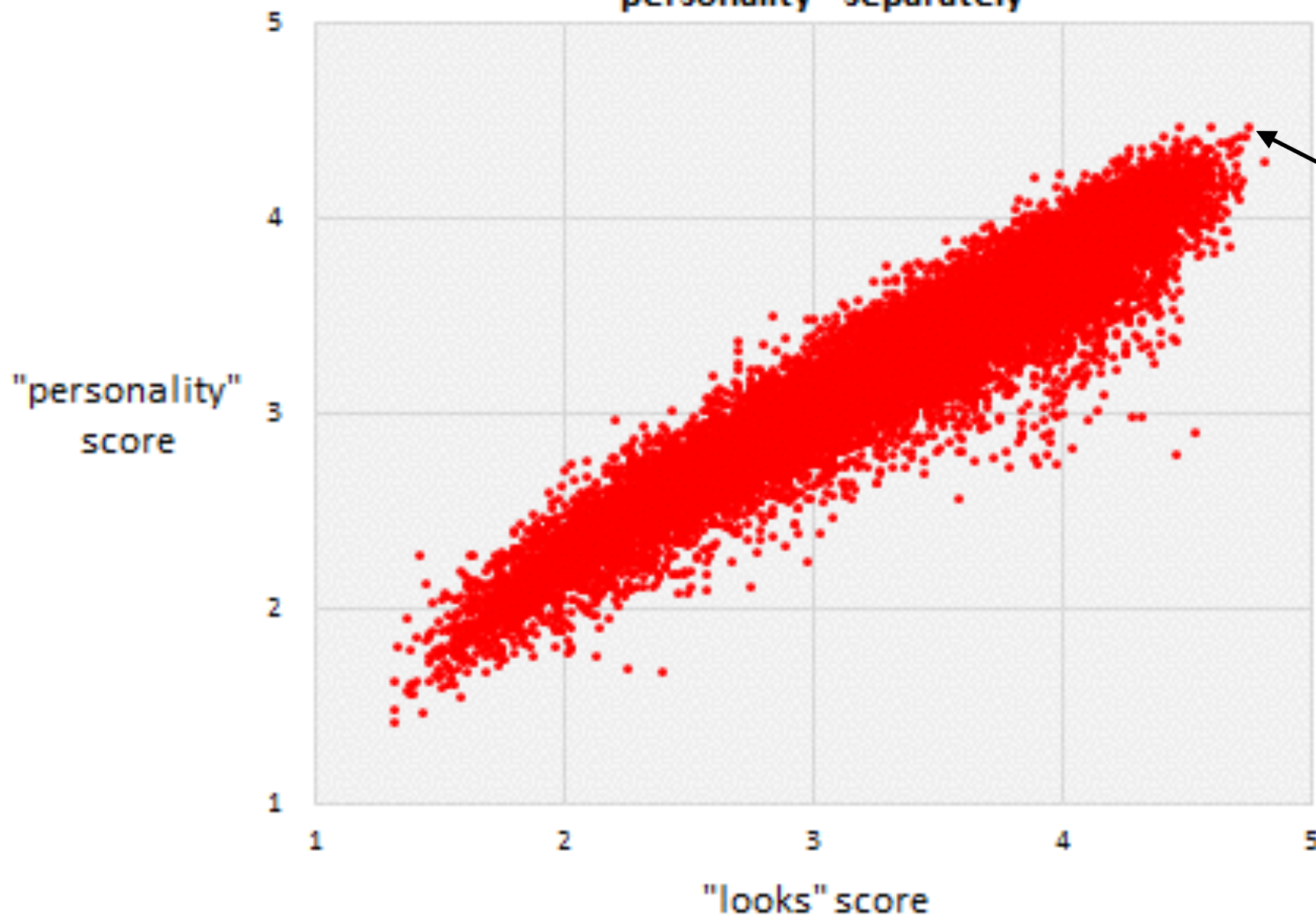
The main navigation bar includes links for MY ACCOUNT, MATCHES, JOURNAL, TESTS, and FUN STUFF. Below this, the user's profile for 'randomcharacter4' is displayed. The profile includes a star icon, a 'Favorites' section, and a 'Stats' section. The 'Stats' section lists various activities: 10 star votes, 5 questions answered, 4 photos, 0 wikiedits accessed, 28 tests taken, 2 stalks per week, 39 messages exchanged, 7 journal comments, and 3 journal posts.

The profile also features a 'QuickMatch!' section with a 'HER FACTS' tab. The facts listed are: AGE 24, LOCATION Denver, Colorado United States, and PERSONA The Window Shopper. A 'SELF-SUMMARY' section follows, containing a paragraph about the user's family, career, and interests.

The profile includes a large photo of a smiling woman, with a 'Loading ...' overlay. To the right of the photo, there are sections for 'Rate her to continue ...' (LOOKS and PERSONALITY) and 'YOU JUST RATED ...' (dbacksfan630). At the bottom right, there is a 'QUICKMATCH PREFERENCES' section with a 'GENDER + ORIENTATION' dropdown menu.

Picture is worth 1000 words?

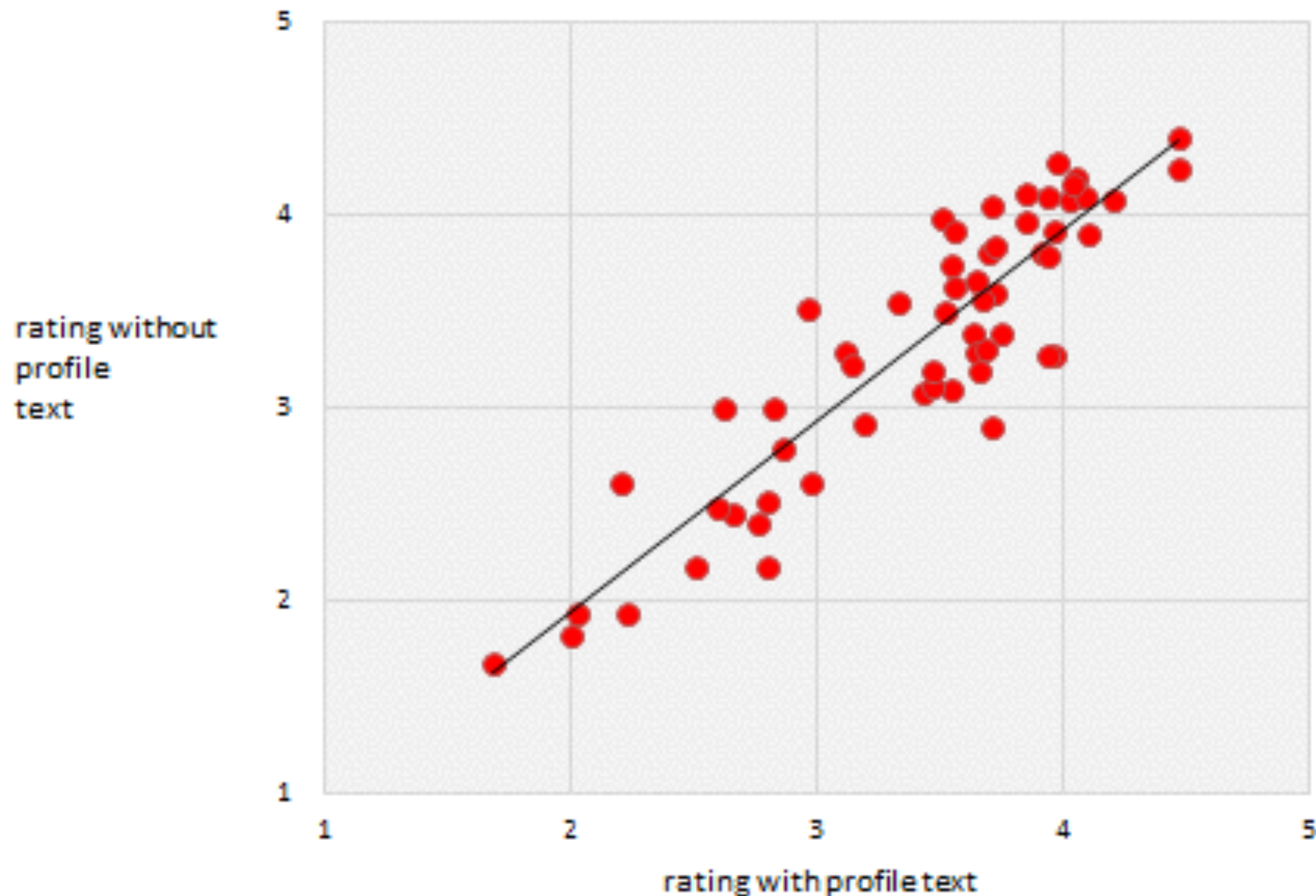
OkCupid ratings back when we let people score "looks" and "personality" separately



Her profile contained **no text**

Picture is worth 1000 words?

people's OkCupid ratings with and without their profile text



Power of Suggestion

For users whose actual match percentage is 30%...
when we DISPLAYED

	<u>30% match</u>	<u>60% match</u>	<u>90% match</u>
Odds of sending one message:	14.2%	16.5%	16.9%

Power of Suggestion

For users whose actual match percentage is 30%...
when we DISPLAYED

	<u>30% match</u>	<u>60% match</u>	<u>90% match</u>
Odds of sending one message:	14.2%	16.5%	16.9%
Given that they sent one message, odds of exchanging four :	9.7%	16.0%	17.4%

Power of Suggestion

Odds of a single message turning into a conversation

		number DISPLAYED to them		
		30% match	60% match	90% match
ACTUAL compatibility of users	30% match	10%	16%	17%
	60% match	13%	13%	16%
	90% match	16%	17%	20%

Good science or creepy
or both?

If you were in charge of
OKCupid, how would you
test the algorithm?