Markets for Data Science ("Crowdsourcing" 2)

April 23, 2019

Data Science CSCI 1951A

Brown University
Instructor: Ellie Pavlick
HTAs: Wennie Zhang, Maulik Dang, Gurnaaz Kaur

*slides borrowed liberally from crowdsourcing-class.org

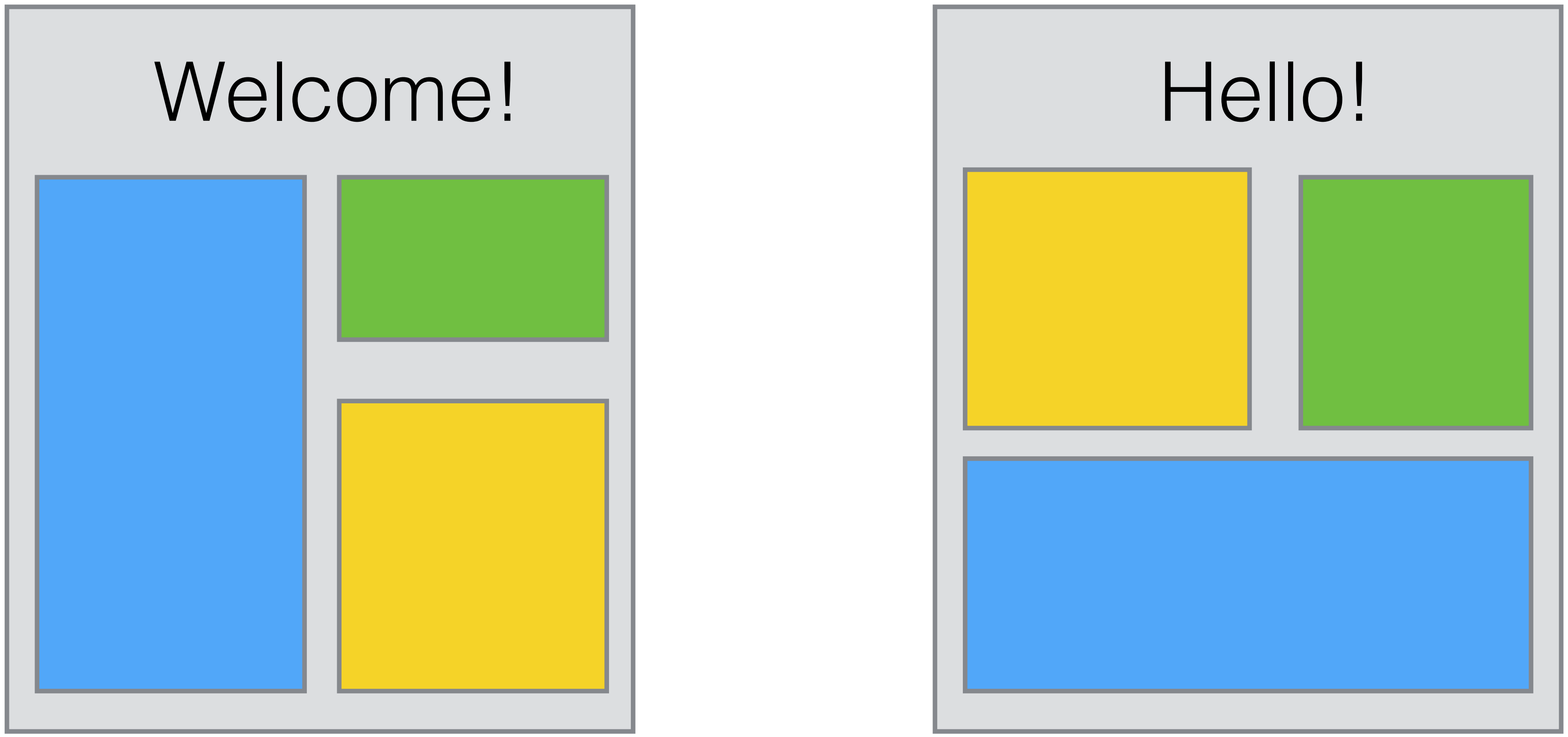
Announcements

Today

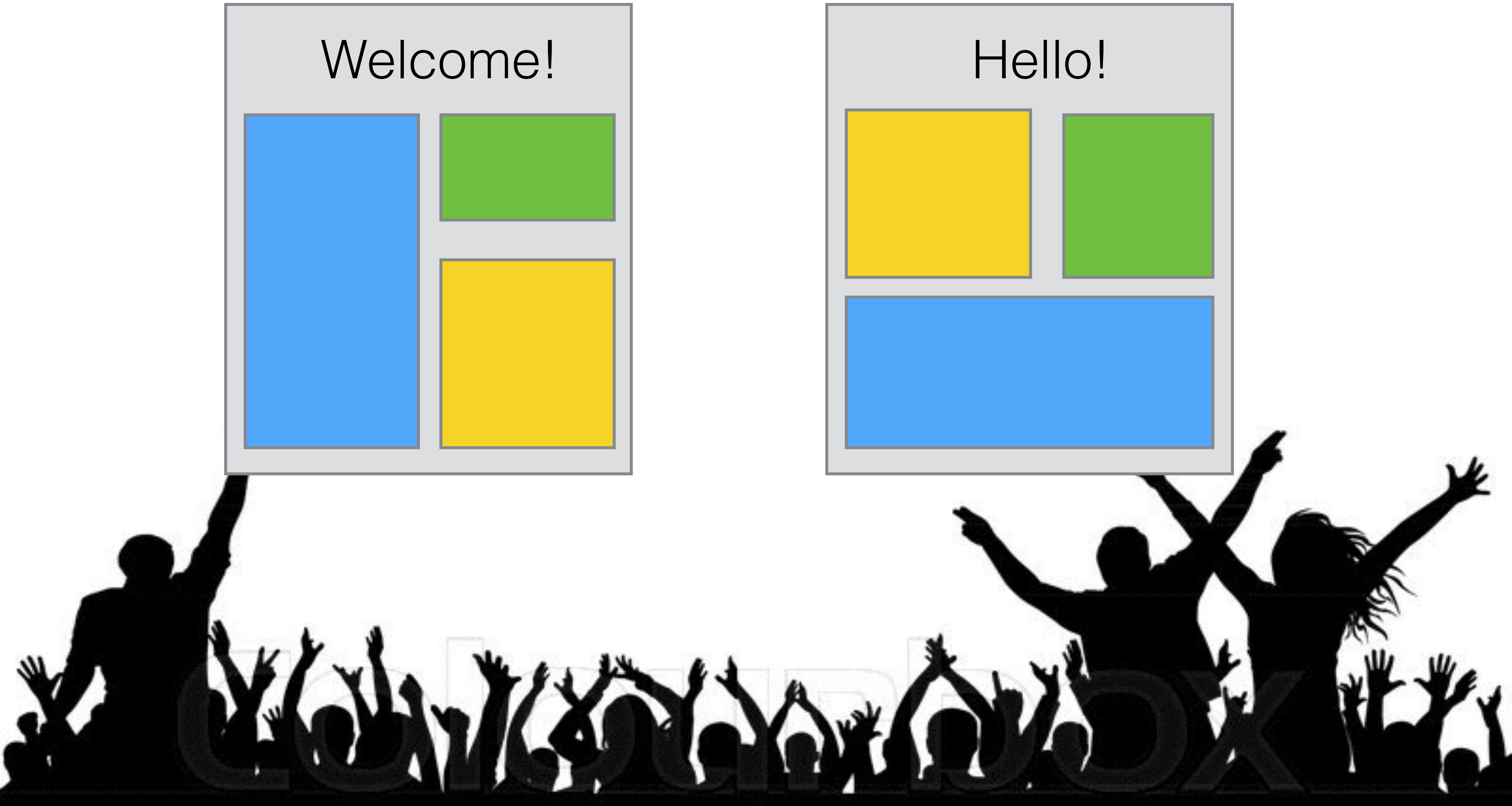
- AB Testing
- Prediction Markets

A/B Testing

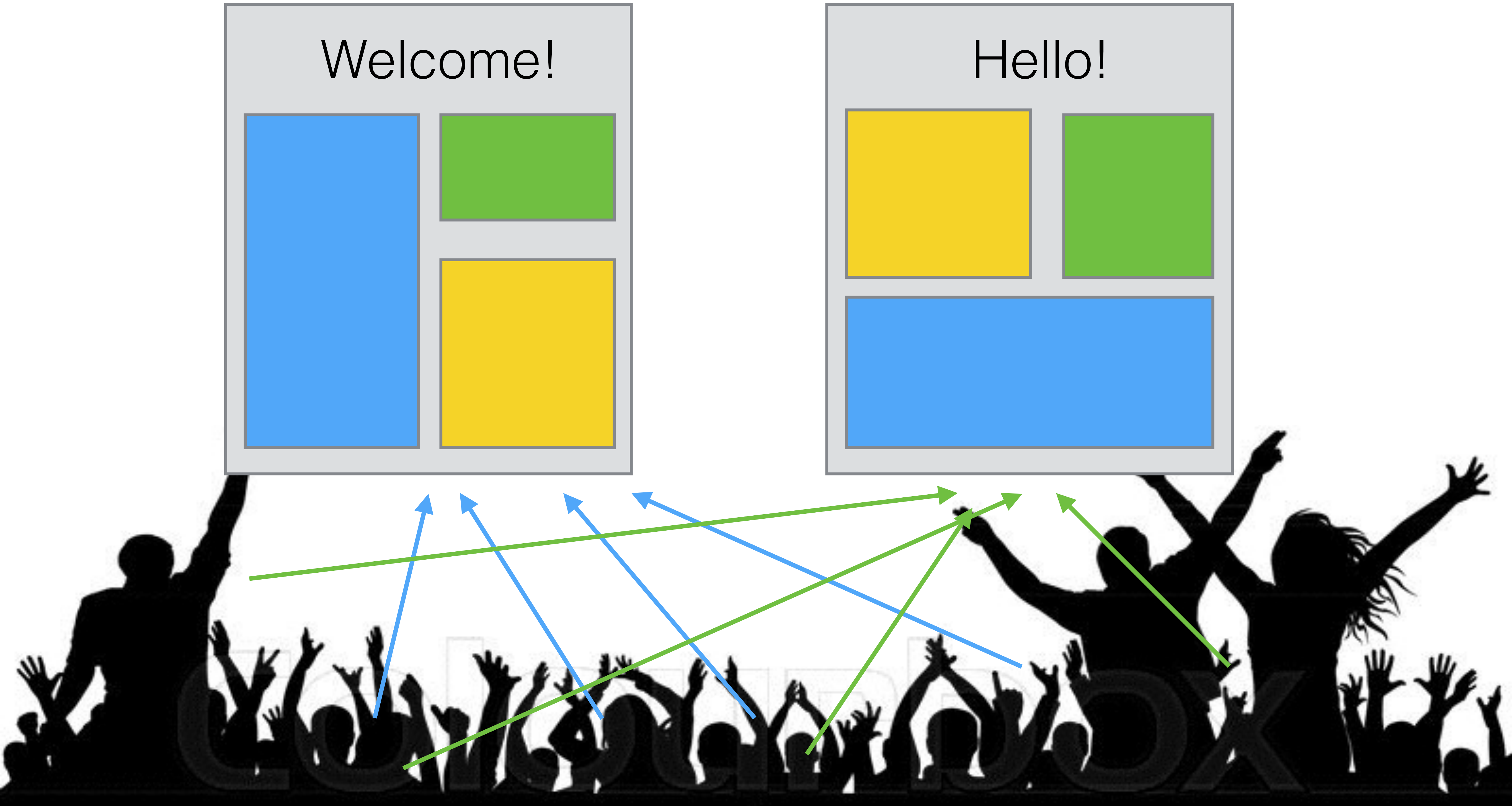
A/B Testing



A/B Testing



A/B Testing



A/B Testing

	Target Outcome?	
	Yes	No
Group A	A_{Yes}	A_{No}
Group B	B_{Yes}	B_{No}

A/B Testing

Chi-Squared Test!

	Target Outcome?	
	Yes	No
Group A	A_{Yes}	A_{No}
Group B	B_{Yes}	B_{No}

$$\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

A/B Testing

Chi-Squared Test!

Target Outcome?

When in doubt, assume proportional...

Group A	A_{Yes}	A_{No}
Group B	B_{Yes}	B_{No}

$$\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

A/B Testing

Observed	Yes	No
Group A	150	50
Group B	6,300	3,500

A/B Testing

Observed	Yes	No
Group A	150	50
Group B	6,300	3,500

Expected	Yes	No
Group A	6,450	
Group B		

A/B Testing

Observed	Yes	No
Group A	150	50
Group B	6,300	3,500

Expected	Yes	No
Group A	6,450/10,000	
Group B		

A/B Testing

Observed	Yes	No
Group A	150	50
Group B	6,300	3,500

Expected	Yes	No
Group A	0.645 x 200	
Group B		

A/B Testing

Observed	Yes	No
Group A	150	50
Group B	6,300	3,500

Expected	Yes	No
Group A	129	71
Group B	6,321	3,479

A/B Testing

Observed

	Yes	No
Group	150	50
Group	6,300	3,500

Expected

	Yes	No
Group	129	71
Group	6,321	3,479

$$\frac{21^2}{129} + \frac{(-21)^2}{71} + \frac{21^2}{6{,}321} + \frac{(-21)^2}{3{,}479}$$

A/B Testing

Observed

	Yes	No
Group	150	50
Group	6,300	3,500

Expected

	Yes	No
Group	129	71
Group	6,321	3,479

3.42 + 6.21 + 0.07 + 0.13 = 9.8

A/B Testing

Observed

	Yes	No
Group	150	50
Group	6,300	3,500

Expected

	Yes	No
Group	129	71
Group	6,321	3,479

$p = 0.002$

$3.42 + 6.21 + 0.07 + 0.13 = 9.8$

What to A/B test?

- Anything manipulatable that effects UX…
  - Web design (color, layout, text)
  - ML models (ads, recommendation systems)
  - Underlying algorithms (“at least not worse” criteria for things like speed ups)

What to A/B test?

- …and that can be quantified by user behavior
    - Whether a user buys your produce
    - Whether they sign up/subscribe to your site
    - Time spent on site

What to A/B test?

- …and that can be quantified by user behavior
  - Whether a user buys your produce
  - Whether they sign up/subscribe to your site
  - Time spent on

Couldnt this start to get..sketchy?

Ethical Principles (IRB)

- Respect for Persons – individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled to protection

- Beneficence – do not harm and maximize possible benefits and minimize possible harms

- Justice – Who ought to receive the benefits of research and bear its burdens?

Legal Guidelines (GDPR)

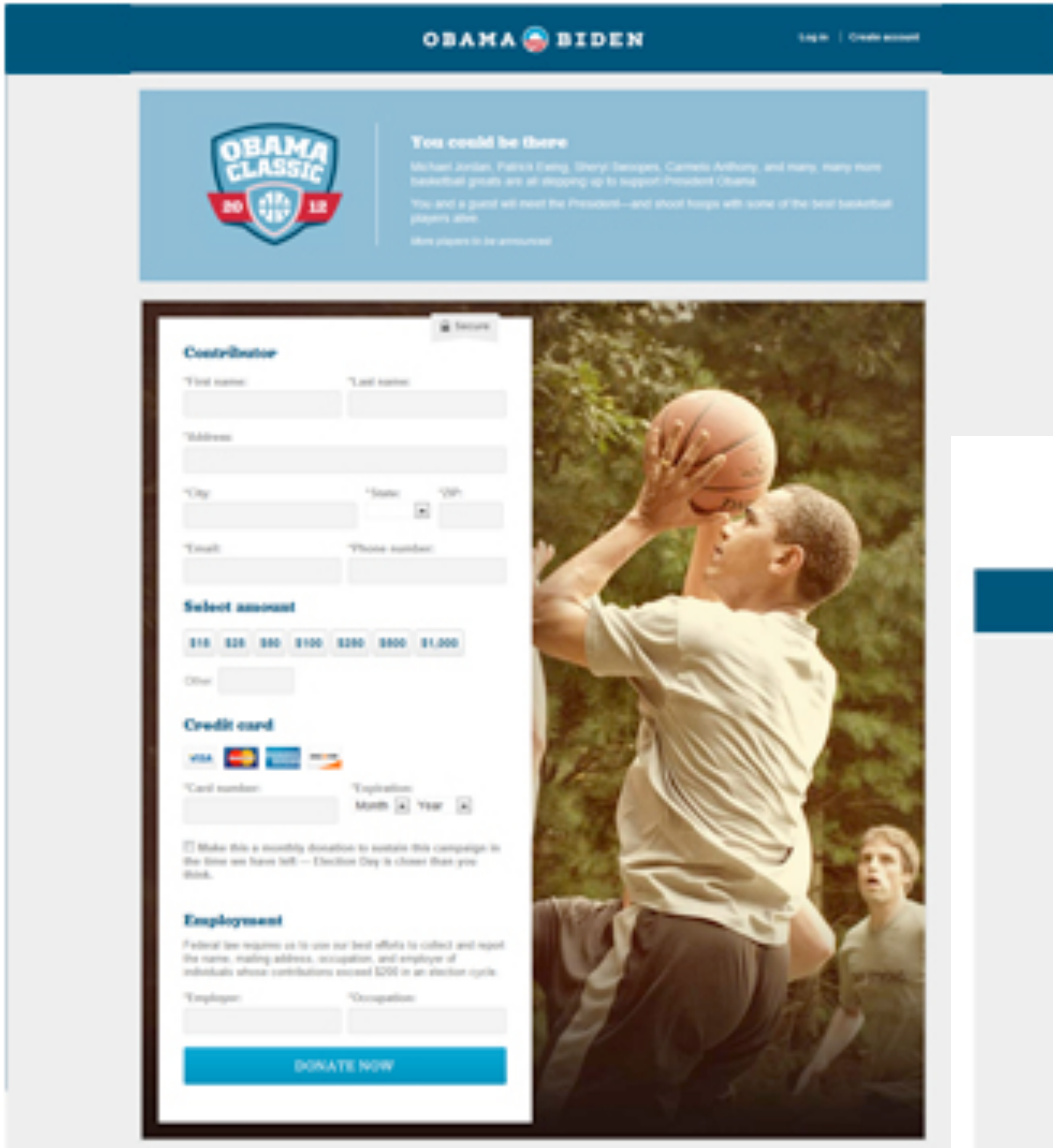
- User Side
  - Information about/access to your personal data
  - Right to be opt-out or have data deleted entirely
  - Request for human-made decisions instead.
- Business Side:
  - Only collect/process data for specific purposes (declare upfront)
  - Can’t use data for purposes other than that for which it was collected
  - Can’t store data longer than necessary

A/B Testing: The Good!

A/B Testing: The Good!

- Kyle Rush, deputy director of front-end 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



+5%

“SEQUENTIAL”

CONTROL

IMAGE VARIATION

+19%

CONTROL

SEGUE COPY +21%

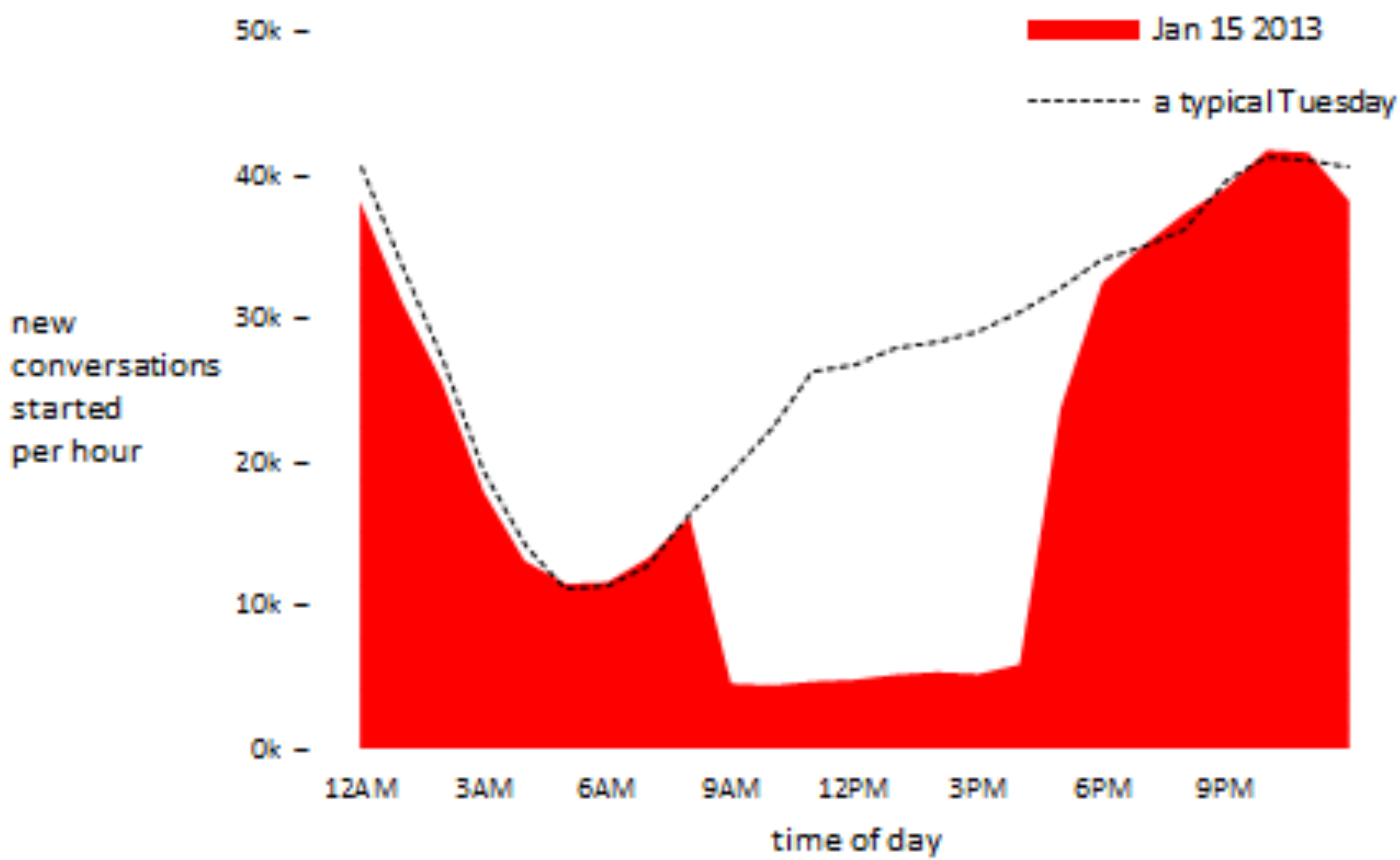
A/B Testing: The Bad…?

We Experiment On Human Beings!

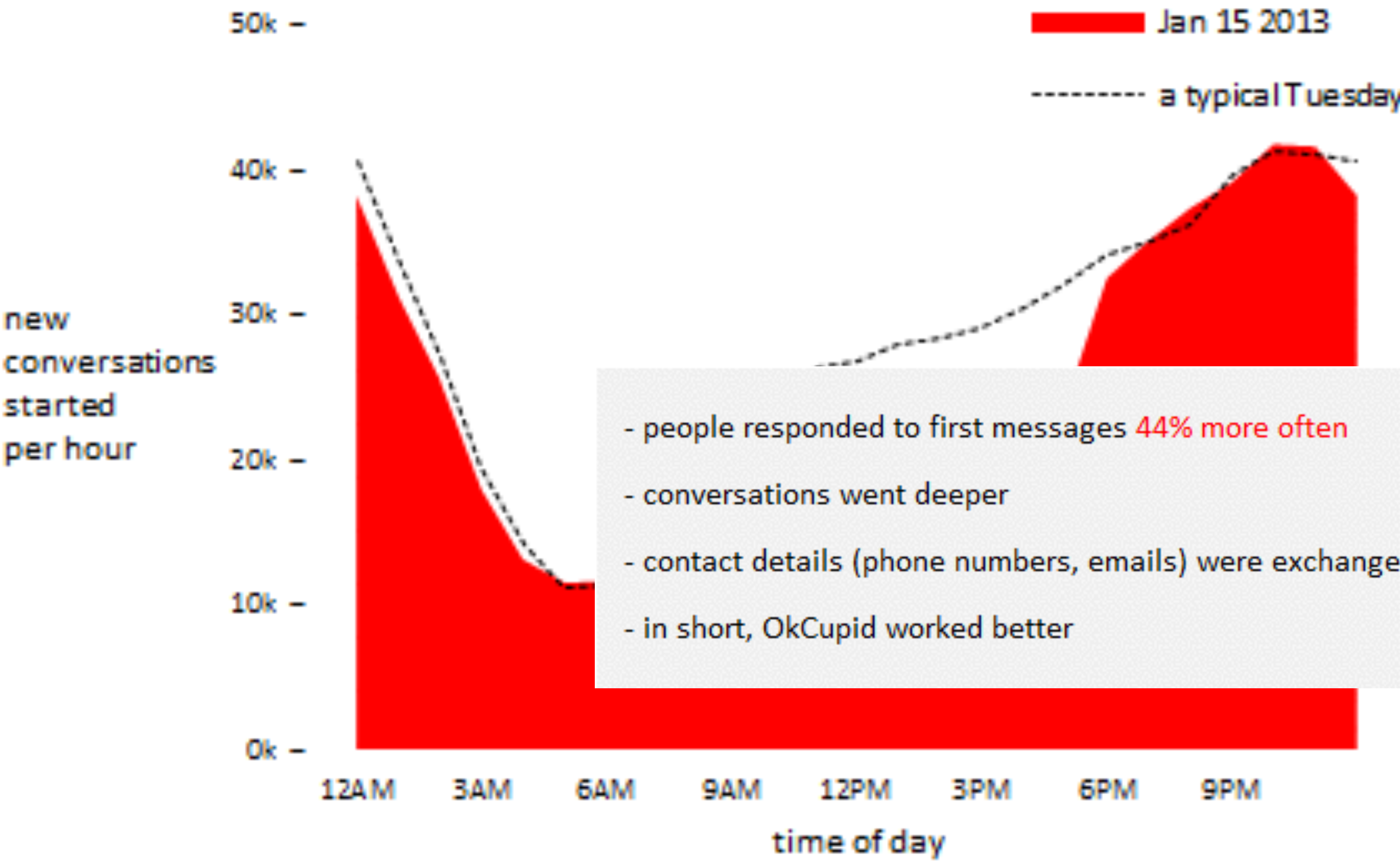
July 28th, 2014 by Christian Rudder

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 young buck, trying to get a potato to cry.

“Love is Blind” Study



“Love is Blind” Study



- 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

Odds of a single message turning into a conversation

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

A/B Testing: The Sketchy

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]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; [b]Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143; and Departments of [c]Communication and [d]Information 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-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. This work also suggests that, in contrast to prevailing assumptions, in-person interaction and nonverbal 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' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experiencer and target.

On Facebook, people frequently express emotions, which are later seen by their friends via Facebook's "News Feed" product (8). Because people's friends frequently produce much more content than one person can view, the News Feed filters posts, stories, and activities undertaken by friends. News Feed is the primary manner by which people see content that friends share. Which content is shown or omitted in the News Feed is determined via a ranking algorithm that Facebook continually develops and tests in the interest of showing viewers 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,

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]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; [b]Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143; and Departments of [c]Communication and [d]Information 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,
and more negative posts;
duced, the opposite patter
emotions expressed by oth
emotions, constituting exp
contagion via social netwo
contrast to prevailing assumptions, in-person interaction and nonverbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.**

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.

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'
hough some shared
late, however, there
oods are contagious
eriencer and target.
motions, which are
ews Feed" product
oduce much more
Feed filters posts,
News Feed is the
that friends share.
News Feed is de-
cebook continually
viewers 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,

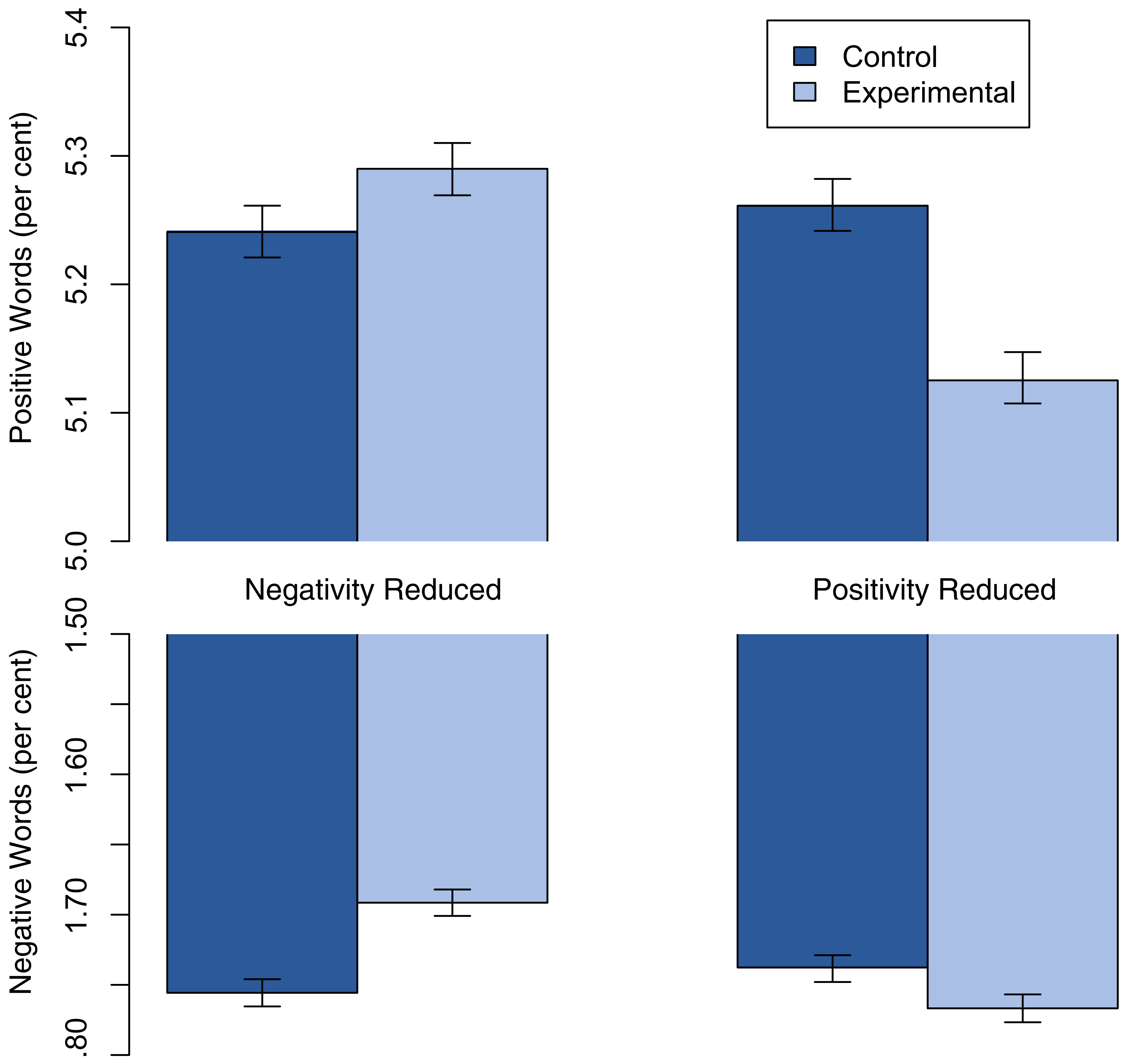


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

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

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."







Prediction Markets

Definitions

- AKA "information market" or "event futures"
- Traders buy/sell contracts which have a payout tied to the unknown outcome of some future event
- Outcomes of events must be unambiguous and verifiable by some predetermined time

Definitions

- Bid/Ask : buyers/sellers chose prices and trades occur only when they match
- Market Makers : individuals agree to make trades, profit from spread

Definitions

- Typical payout is like in horse racing - all money is pooled and then divided among winners
- Incentive scheme can be real or virtual/play money

Table 1: Contract Types—Estimating Uncertain Quantities or Probabilities

Contract	Example	Details	Reveals market expectation of…
Winner-takes-all	Event y: Al Gore wins the popular vote	Contract costs $\$p$ Pays \$1 if and only if event y occurs Bid according to value of $\$p$	Probability that event y occurs, $p(y)$
Index	Contract pays \$1 for every percentage point of the popular vote won by Al Gore	Contract pays $\$y$.	Mean value of outcome y: $E[y]$
Spread	Contract pays even money if Gore wins more than y^*% of the popular vote.	Contract costs \$1 Pays \$2 if $y>y^*$ Pays \$0 otherwise. Bid according to the value of y^*.	Median value of y.

Theory

- Prices should be (and often are) **efficient**: price should be equal to expected payout. (Although small markets may absorb information less quickly than larger markets.)

- Marginal trades should be (and often are) **rational**: no systematic biases should arise. (Although people often trade according to desires rather than beliefs.)

- Markets should (and often do) contain **few arbitrage opportunities**: the same contracts should trade at the same price on different exchanges

Market A sells "Obama wins" contract for $0.75
Market B sells "Obama wins" contract for $0.50

You have not a penny to your name	\$0	\$0
You short sell 100 contracts on A. (I.e. you borrow contracts and sell them. You will have to return them later.)	+\$75	\$75
You buy 100 contracts in market B	-\$50	\$25

OBAMA WINS!!

M [illegible] a wins" contract for $0.75

[illegible] Obama wins" contract for $0.50

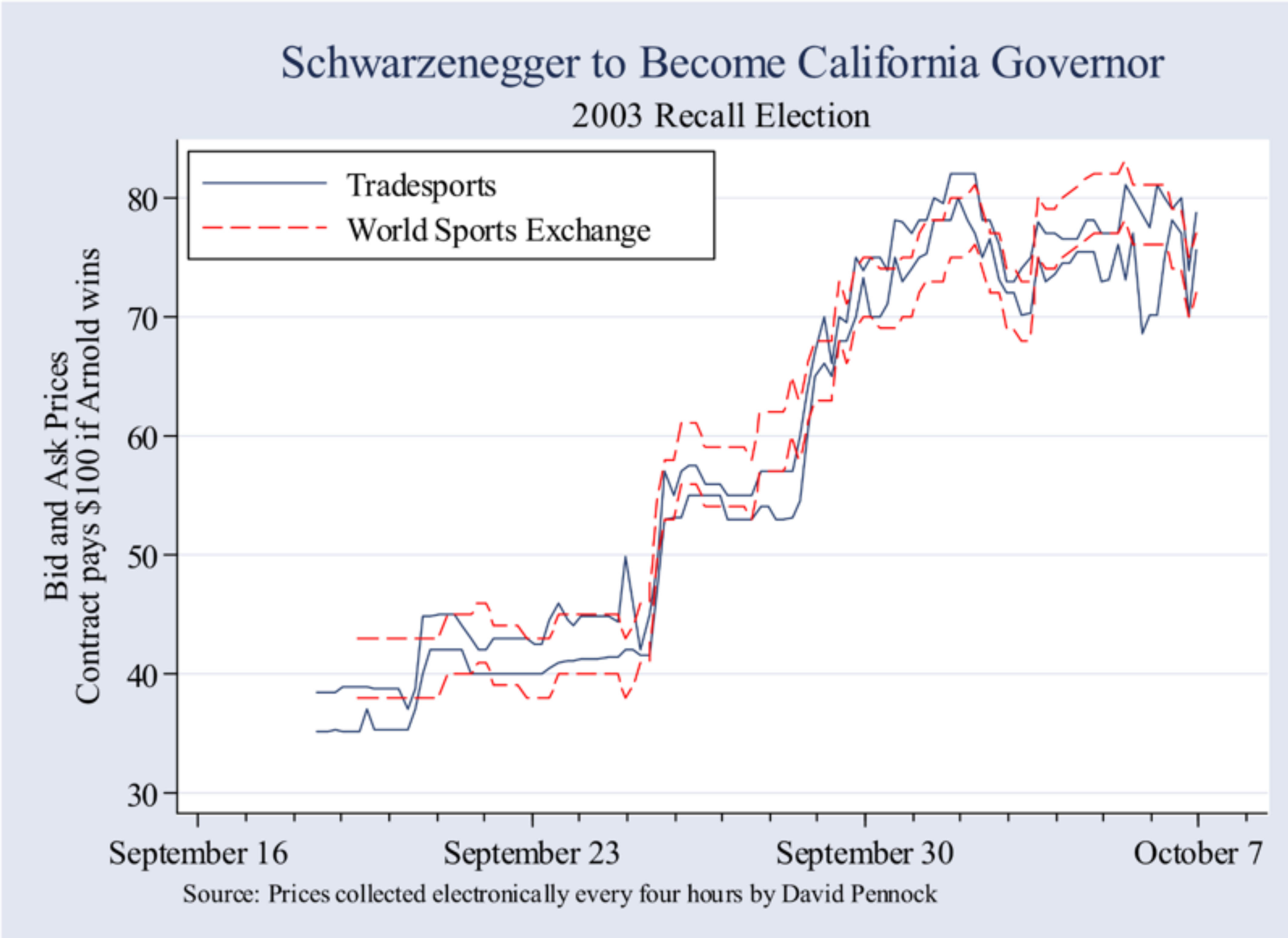
You have not a penny to your name	$0	$0
You short sell 100 contracts on A. (I.e. you borrow contracts and sell them. You will have to return them later.)	+$75	$75
You buy 100 contracts in market B	-$50	$25
Your contracts on market B are worth $100.	+$100	$125
You return 100 shares that you borrowed on Market A (now worth $100).	-$100	$25
Profit		$25

OBAMA LOSES!!

M… …a wins" contract for $0.75

…Obama wins" contract for $0.50

You have not a penny to your name	$0	$0
You short sell 100 contracts on A. (I.e. you borrow contracts and sell them. You will have to return them later.)	+$75	$75
You buy 100 contracts in market B	-$50	$25
Your contracts on market B are worth $0.	+$0	$25
You return 100 shares that you borrowed on Market A (now worth $0).	$0	$25
Profit		$25


Schwarzenegger to Become California Governor
2003 Recall Election
Tradesports
World Sports Exchange
Bid and Ask Prices
Contract pays $100 if Arnold wins
80
70
60
50
40
30
September 16
September 23
September 30
October 7
Source: Prices collected electronically every four hours by David Pennock

Theory

- For simplicity, our definition of prediction markets :
  - Does not include markets where holding the good is inherently enjoyable (e.g. sports betting)
  - Does not include markets large enough to allow risk sharing
  - Includes only risk neutral probabilities
- (As always, these assumptions can be relaxed...)

Theory

- Binary contracts paying $1 dollar if event occurs, $0 otherwise
- Wealth is orthogonal to event outcome and to beliefs
- Market is large, and participants are “price takers”
- Log utility
- Beliefs are heterogeneous and reflect private, noisy signals of whether the event will occur

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j\pi]$$

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

y = the person's wealth

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + \boxed{x_j}(1-\pi)] + (1-q_j) Log[y - \boxed{x_j}\pi]$$

y = the person's wealth

x_j = the number of contracts person j should buy

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j)Log[y - x_j\pi]$$

y = the person's wealth
x_j = the number of contracts person j should buy
∏ = prince of the contract

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

y = the person's wealth

x_j = the number of contracts person j should buy

∏ = prince of the contract

q_j = person j's believed probability of the event

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j)Log[y - x_j\pi]$$

P(winning)

y = the person's wealth

x_j = the number of contracts person j should buy

∏ = prince of the contract

q_j = person j's believed probability of the event

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

P(winning) x (wealth if you win)

y = the person's wealth

x_j = the number of contracts person j should buy

∏ = prince of the contract

q_j = person j's believed probability of the event

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

P(winning) x (wealth if you win)
+ P(Losing)

y = the person's wealth
x_j = the number of contracts person j should buy
∏ = prince of the contract
q_j = person j's believed probability of the event

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

P(winning) x (wealth if you win)
+ P(losing) x (wealth if you lose)

y = the person's wealth
x_j = the number of contracts person j should buy
Π = prince of the contract
q_j = person j's believed probability of the event

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

$$x_j^* = y \frac{q_j - \pi}{\pi(1-\pi)}$$

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j\pi]$$

$$x_j^* = y\frac{q_j - \pi}{\pi(1-\pi)}$$

Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j\pi]$$

$$x_j^* = y\frac{q_j - \pi}{\pi(1-\pi)}$$

demand is 0 when price = beliefs

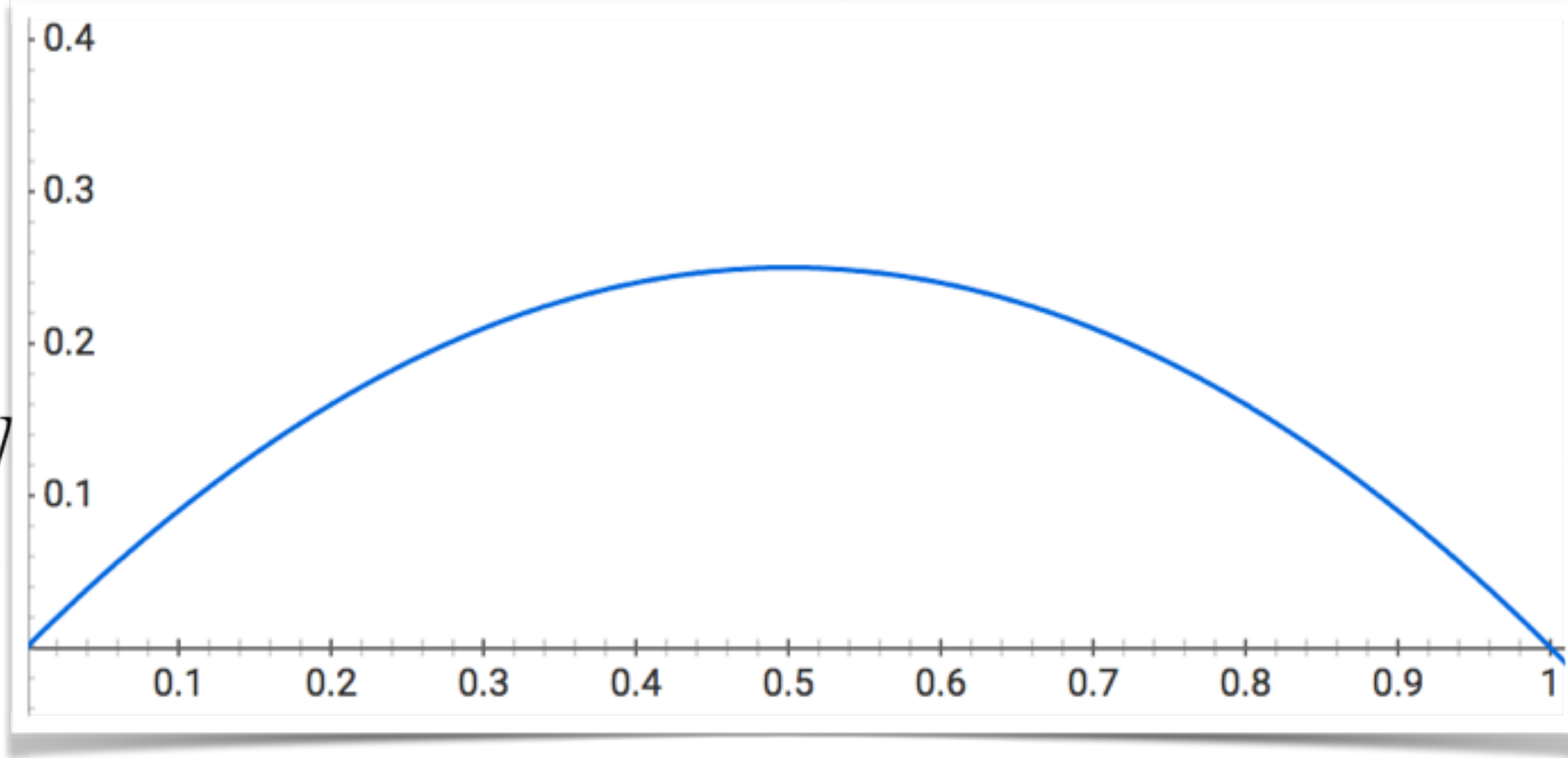
Theory

$$\underset{\{x\}}{Max}\ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

$$x_j^* = y \frac{q_j - \pi}{\pi(1-\pi)}$$

demand is "linear in beliefs"
(given y, demand increases with q)

$$\underset{\{x\}}{Max}\ EU \quad \cdots \quad - x_j \pi]$$



$$x_j^* = y \frac{q_j - \pi}{\pi(1-\pi)}$$

demand is "decreasing in risk"
(demand lower when pi is close to 0.5)

Price equal to mean(q) when supply = demand

$$\int_{-\infty}^{\pi} y\frac{q-\pi}{\pi(1-\pi)}f(q)dq = \int_{\pi}^{\infty} y\frac{\pi-q}{\pi(1-\pi)}f(q)dq$$

$$\frac{y}{\pi(1-\pi)}\int_{-\infty}^{\pi}(q-\pi)f(q)dq = \frac{y}{\pi(1-\pi)}\int_{\pi}^{\infty}(\pi-q)f(q)dq$$

$$\pi = \int_{-\infty}^{\infty} qf(q)dq = \overline{q}$$

Price equal to mean(q) when

$$\int_{-\infty}^{\pi} y \frac{q-\pi}{\pi(1-\pi)} f(q)dq =$$

At any price below equilibrium, consumers will be better off than producers (they are getting away with paying too little).

$$\frac{y}{\pi(1-\pi)} \int_{-\infty}^{\pi} (q-\pi) f(q)dq = \frac{y}{\pi(1-\pi)} \int_{\pi}^{\infty} (\pi-q) f(q)dq$$

$$\pi = \int_{-\infty}^{\infty} qf(q)dq = \overline{q}$$

Price equal to mean(q) when supply = demand

At any price above equilibrium, producers will be better off than consumers (they are getting away with charging too much).

$$= \int_{\pi}^{\infty} y \frac{\pi - q}{\pi(1-\pi)} f(q)dq$$

$$\frac{y}{\pi(1-\pi)} \int_{-\infty}^{\pi} (q-\pi) f(q)dq = \frac{y}{\pi(1-\pi)} \int_{\pi}^{\infty} (\pi - q) f(q)dq$$

$$\pi = \int_{-\infty}^{\infty} qf(q)dq = \overline{q}$$

Price equal to mean(q) when supply = demand

All the well-off-ness of consumers $$= \int_{\pi}^{\infty} y \frac{\pi - q}{\pi(1-\pi)} f(q)\,dq$$

$$\frac{y}{\pi(1-\pi)} \int_{-\infty}^{\pi} (q-\pi) f(q)\,dq = \frac{y}{\pi(1-\pi)} \int_{\pi}^{\infty} (\pi - q) f(q)\,dq$$

$$\pi = \int_{-\infty}^{\infty} q f(q)\,dq = \overline{q}$$

Price equal to mean(q) when supply = demand

All the well-off-ness of consumers = All the well-off-ness of producers

$$\frac{y}{\pi(1-\pi)}\int_{-\infty}^{\pi}(q-\pi)f(q)dq = \frac{y}{\pi(1-\pi)}\int_{\pi}^{\infty}(\pi-q)f(q)dq$$

$$\pi = \int_{-\infty}^{\infty} qf(q)dq = \overline{q}$$

Price equal to mean(q) when supply = demand

All the well-off-ness of consumers = All the well-off-ness of producers

Math

$$\pi = \int_{-\infty}^{\infty} q f(q) dq = \overline{q}$$

Price equal to mean(q) when supply = demand

All the well-off-ness of consumers = All the well-off-ness of producers

Math

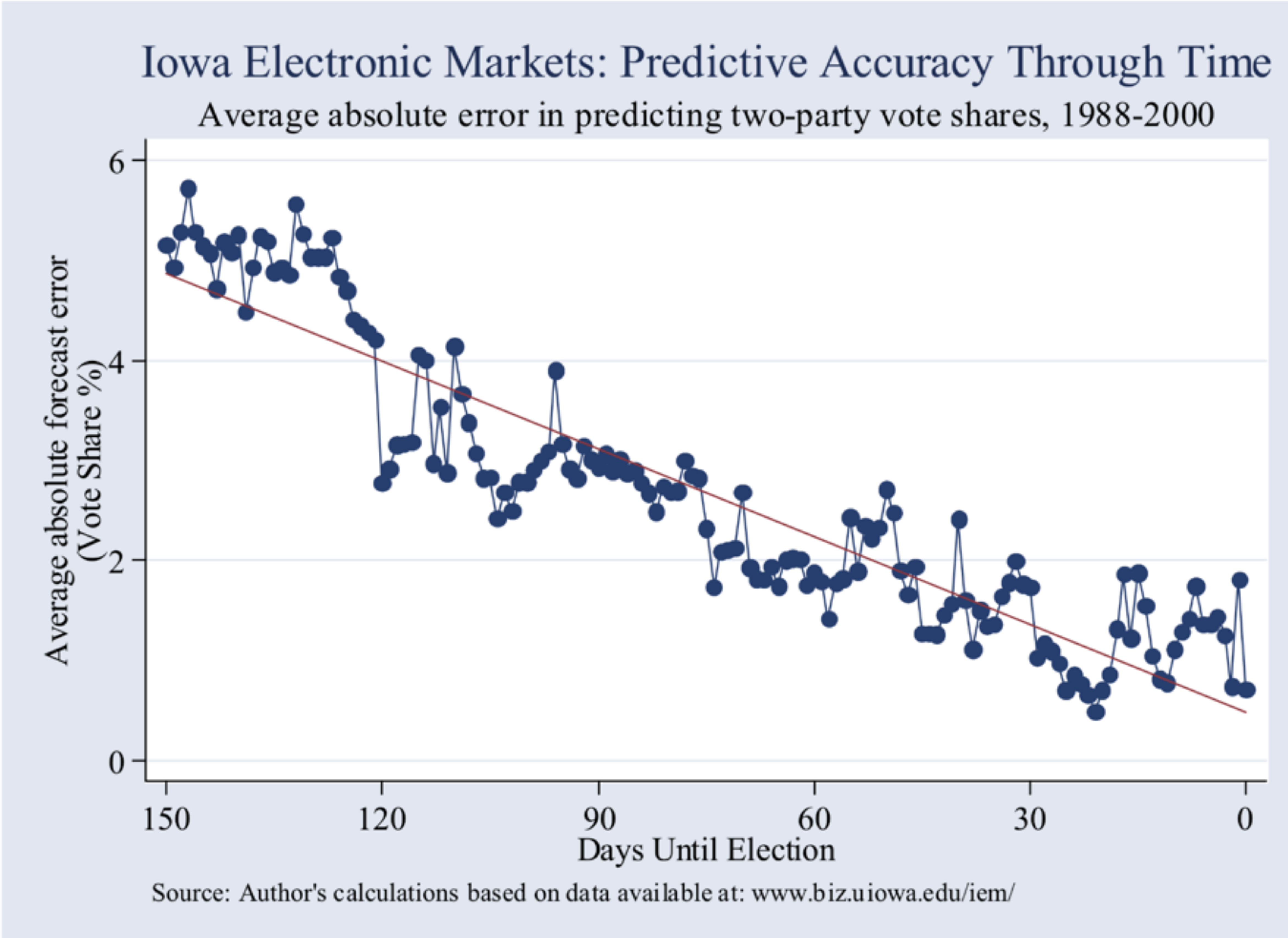
$\pi =$ Average of all participants beliefs

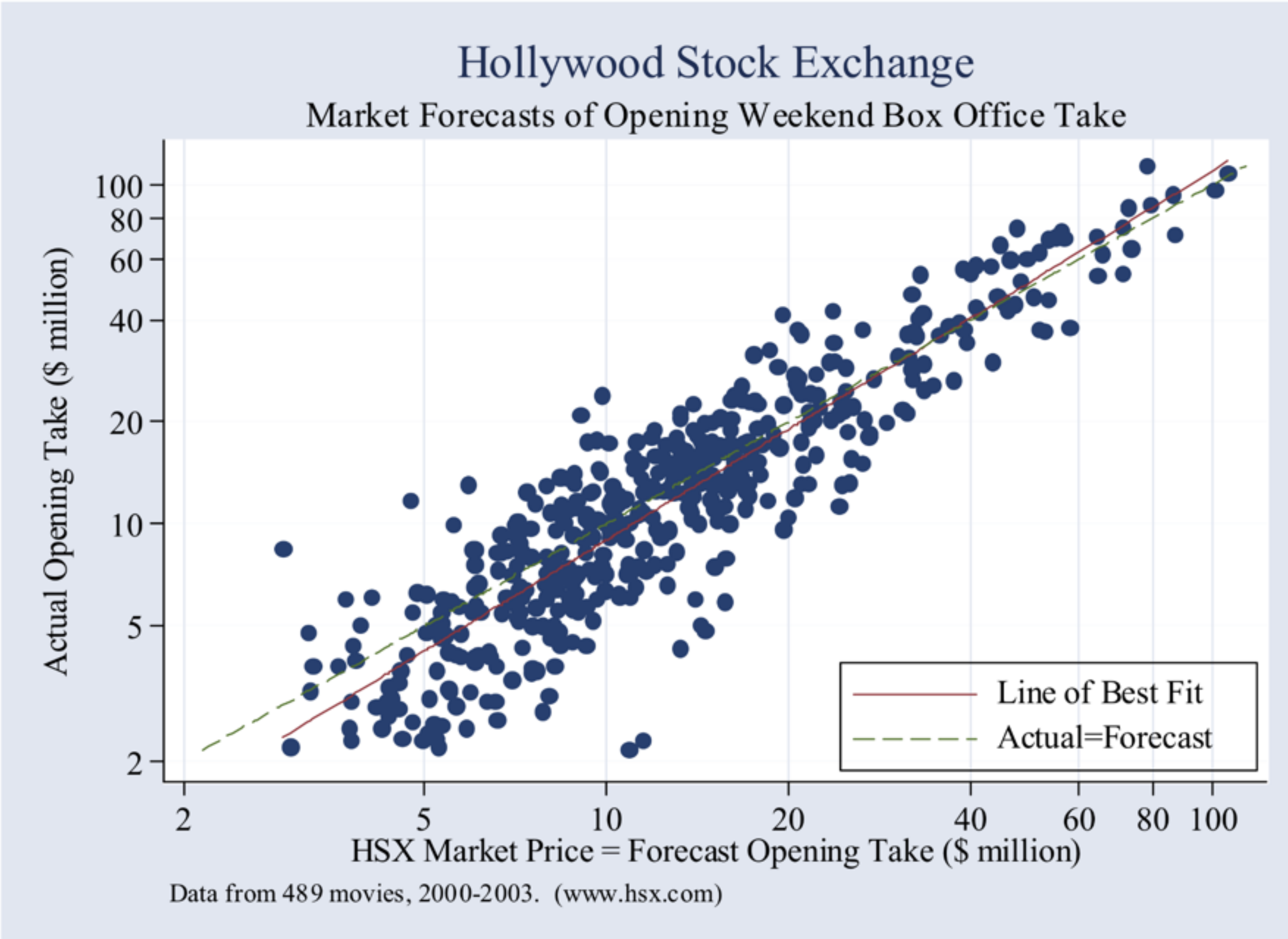
Practice

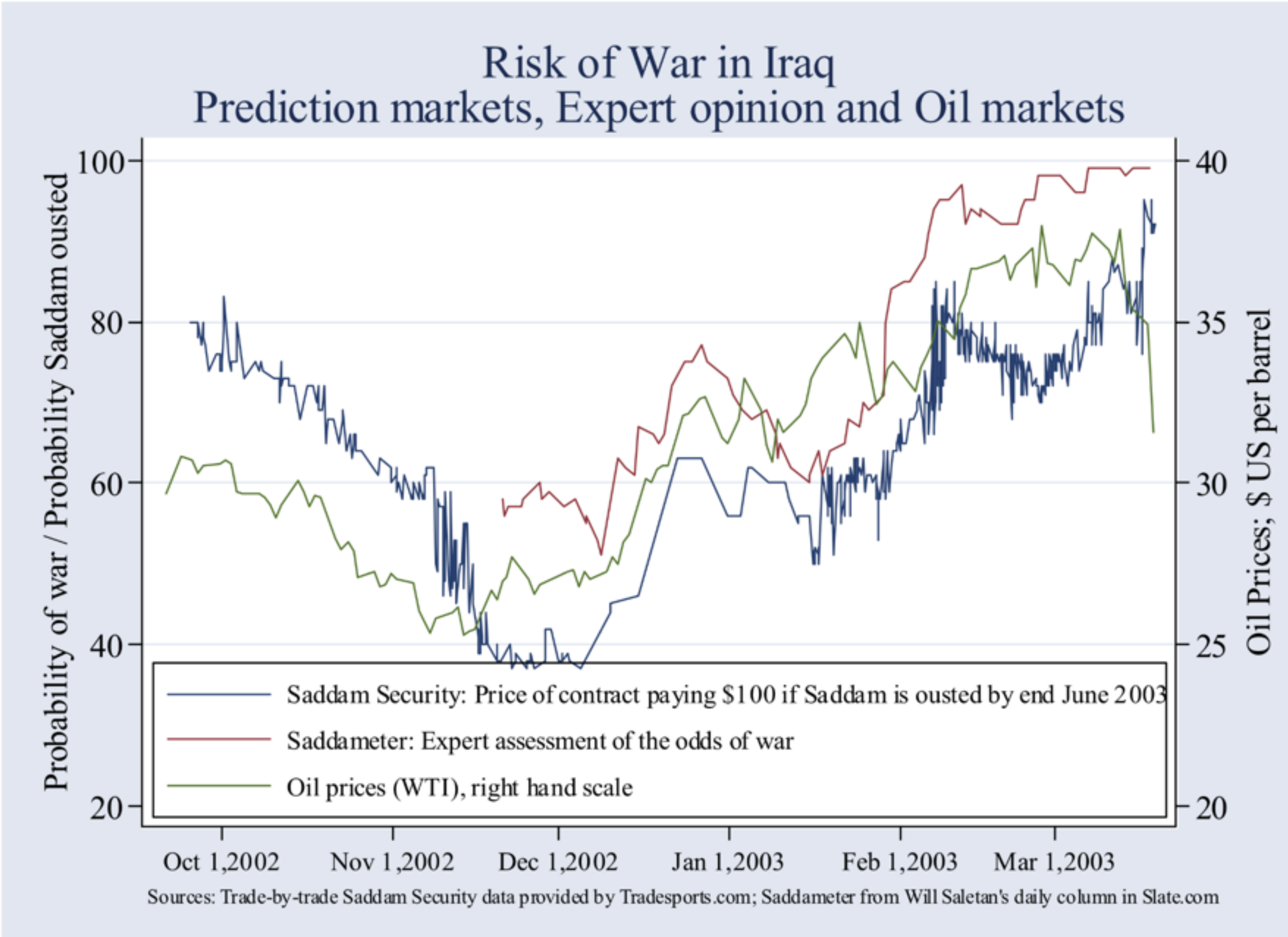
- For business/fun : Intrade, Tradesports
- For research : Iowa Election Markets
- For government : PAM
- For companies internally: HP (printer sales), Siemens (ability to meet deadlines)

Table 2: Prediction Markets

Market	Focus	Typical turnover on an event ($US)
Iowa Electronic Markets <www.biz.iowa.edu/iem> *Run by University of Iowa*	Small-scale election markets. Similar markets are run by: UBC (Canada) <www.esm.buc.ca> and TUW (Austria) <http://ebweb.tuwien.ac.at/apsm/>	Tens of thousands of dollars (Traders limited to $500 positions)
TradeSports <www.tradesports.com> *For profit company*	Trade in a rich set of political futures, financial contracts, current events, sports and entertainment	Hundreds of thousands of dollars
Economic Derivatives <www.economicderivatives.com> *Run by Goldman Sachs and Deutsche Bank*	Large-scale financial market trading in the likely outcome of future economic data releases	Hundreds of millions of dollars
Newsfutures <www.newsfutures.com> *For profit company*	Political, finance, current events and sports markets. Also technology and pharmaceutical futures for specific clients.	Virtual currency redeemable for monthly prizes (such as a TV)
Foresight Exchange <www.ideosphere.com> *Non-profit research group*	Political, finance, current events, science and technology events suggested by clients.	Virtual currency
Hollywood Stock Exchange <www.hsx.com> *Owned by Cantor Fitzgerald*	Success of movies, movie stars, awards, including a related set of complex derivatives and futures. Data used for market research.	Virtual currency.


Iowa Electronic Markets: Predictive Accuracy Through Time
Average absolute error in predicting two-party vote shares, 1988-2000
Average absolute forecast error
(Vote Share %)
6
4
2
0
150
120
90
60
30
0
Days Until Election
Source: Author's calculations based on data available at: www.biz.uiowa.edu/iem/


Hollywood Stock Exchange
Market Forecasts of Opening Weekend Box Office Take
Actual Opening Take ($ million)
100
80
60
40
20
10
5
2
2
5
10
20
40
60
80
100
HSX Market Price = Forecast Opening Take ($ million)
Line of Best Fit
Actual=Forecast
Data from 489 movies, 2000-2003. (www.hsx.com)



Risk of War in Iraq
Prediction markets, Expert opinion and Oil markets
Probability of war / Probability Saddam ousted
100
80
60
40
20
40
35
30
25
20
Oil Prices; $ US per barrel
Saddam Security: Price of contract paying $100 if Saddam is ousted by end June 2003
Saddameter: Expert assessment of the odds of war
Oil prices (WTI), right hand scale
Oct 1,2002
Nov 1,2002
Dec 1,2002
Jan 1,2003
Feb 1,2003
Mar 1,2003
Sources: Trade-by-trade Saddam Security data provided by Tradesports.com; Saddameter from Will Saletan's daily column in Slate.com

Google's Prediction Market

http://www.eecs.harvard.edu/cs286r/courses/fall10/papers/GooglePredictionMarketPaper.pdf

Research Questions

"...internal prediction can provide insight into how organizations process information. Prediction markets provide employees with **incentives for truthful revelation** and can capture changes in opinion at a **much higher frequency than surveys**, allowing one to track how information moves around an organization and how it responds to external events."

Cowgill, Wolfers, and Zitzewitz 2009

Research Questions

- **Optimism** in entrepreneurial firms: "Entrepreneur’s curse" suggests that entrepreneurial firms tend to be optimistically biased about their potential for success.

- Employee **communication** in organization: Firms pay high costs to cluster in places like Silicon Valley; prediction markets can be used as high-frequency, market-incentivized surveys to track information flows in real-time.

- Social networks and **information flows** among investors: Prediction markets as a way to test the importance of physical proximity and social networks in facilitating information sharing.

Market Overview

- Launched April 2005, each quarter from 2005Q2 to 2007Q3 had 25-30 markets
- Question that has 2-5 mutually exclusive and exhaustive answers, e.g.
  - Q: “How many users will Gmail have?”
  - A : “Fewer than X users”, “Between X and Y”, “More than Y”.
- Answer corresponds to a security that is worth one “Gooble” if the answer turns out to be correct. At the end of the quarter, Goobles were converted into raffle tickets and prizes were raffled off.
- Prize budget was $10,000 per quarter ($25-100 per trader)
- Out of 6,425 employees who had accounts, 1,463 placed at least one trade.

Market Overview

- Short selling is not allowed; traders can buy a set of securities and then sell the ones they choose.
- There is no automated market maker, but several employees did create robotic traders that sometimes played this role.
- Volume in “fun” and “serious” markets are positively correlated

Market Overview

- Participants were not representative of Google as a whole
- More likely to be in programming roles
- More likely to be in Mountain View or New York campuses
- More quantitative backgrounds (e.g. undergraduate major)
- More interest in investing or poker (e.g. mailing lists)
- Employed longer, less likely to leave after study
- Slightly more senior (levels from CEO)

Market Overview

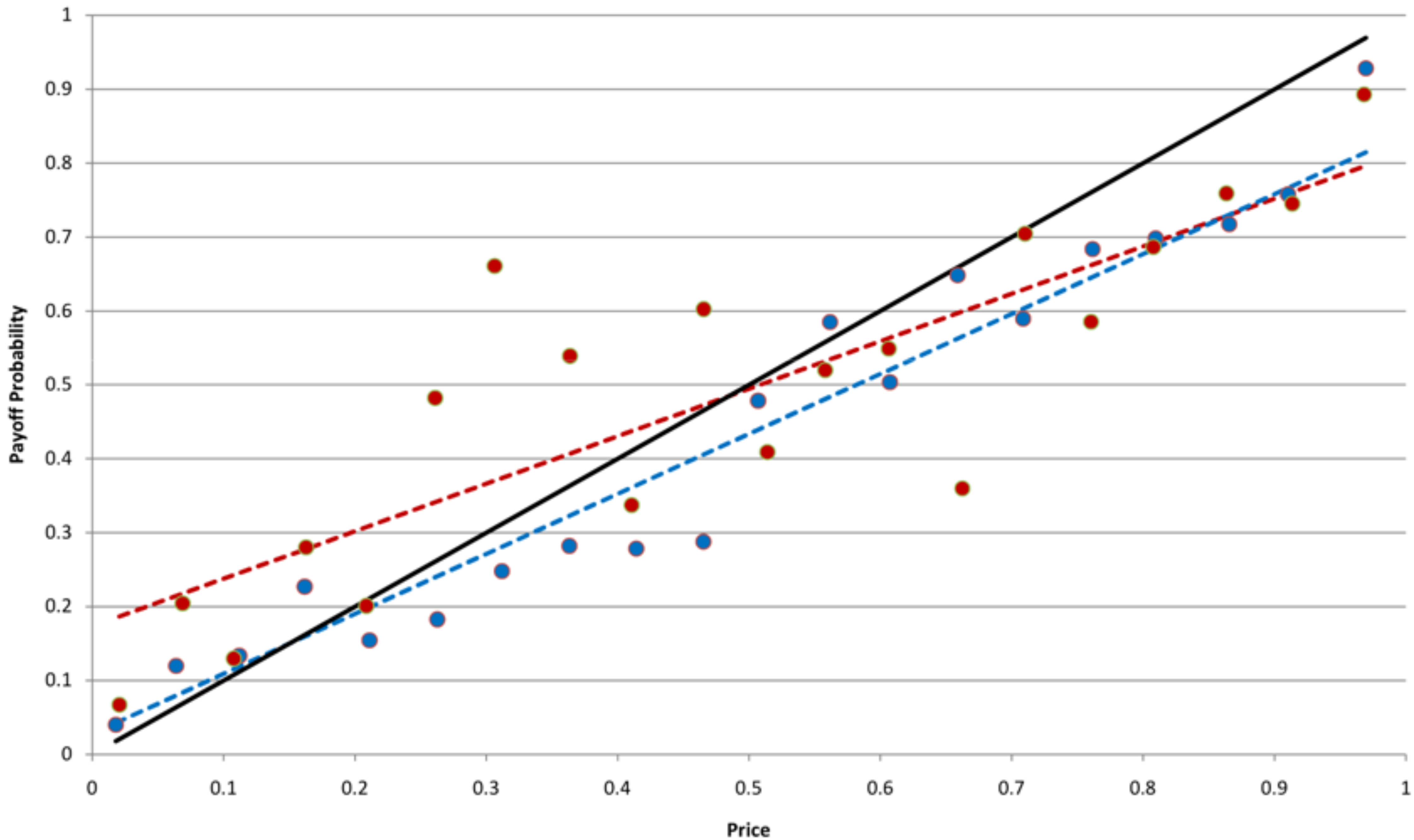
Table 1. Prediction markets at Google

Type	Example	Share of markets
Demand forecasting	# of Gmail users at end of quarter	20%
Performance	Google Talk quality rating	15%
Company news	Russia office to open	10%
Industry news	Will Apple release an Intel-based Mac?	19%
Decision markets	Will users of feature A users use feature B more?	2%
Fun	How many "rotten tomatoes" will Episode III get?	33%
Unique participants		1,463
Orders		253,192
Trades		70,706
Markets run (questions)		270
Securities (answers)		1,116

Biases

- Overpricing of favorites
- Underpricing of extreme outcomes
- Short aversion
- Optimism

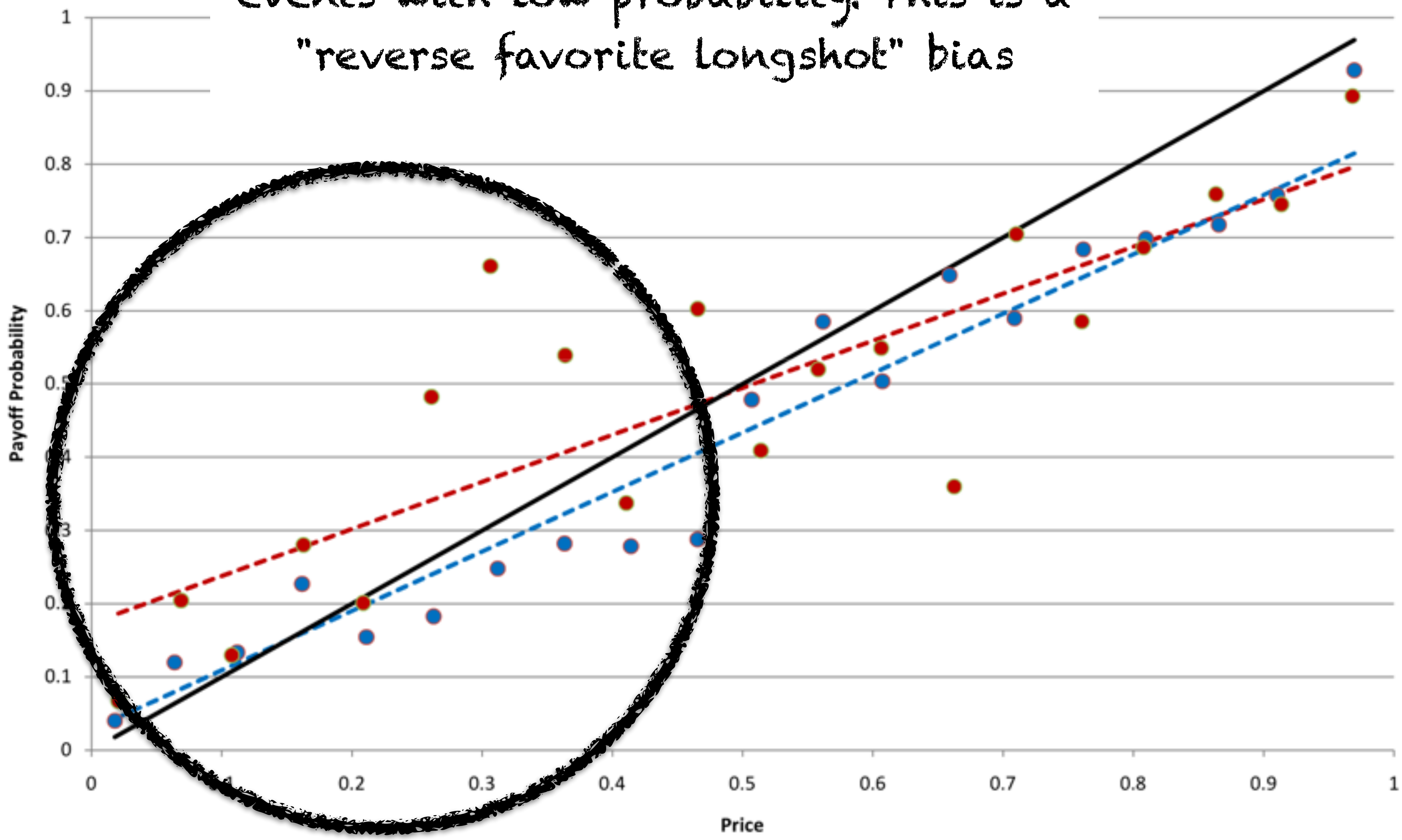
Figure 2. Prices and Probabilities in Two and Five-outcome Markets



Trades in two (red) and five-outcome (blue) markets (22,452and 42,416, respectively) are sorted into 20 bins according to price (i.e., 0-5, 5-10, etc.), and then average price and payoff probability for the bin is plotted. Dashed lines plot regression equations using OLS.

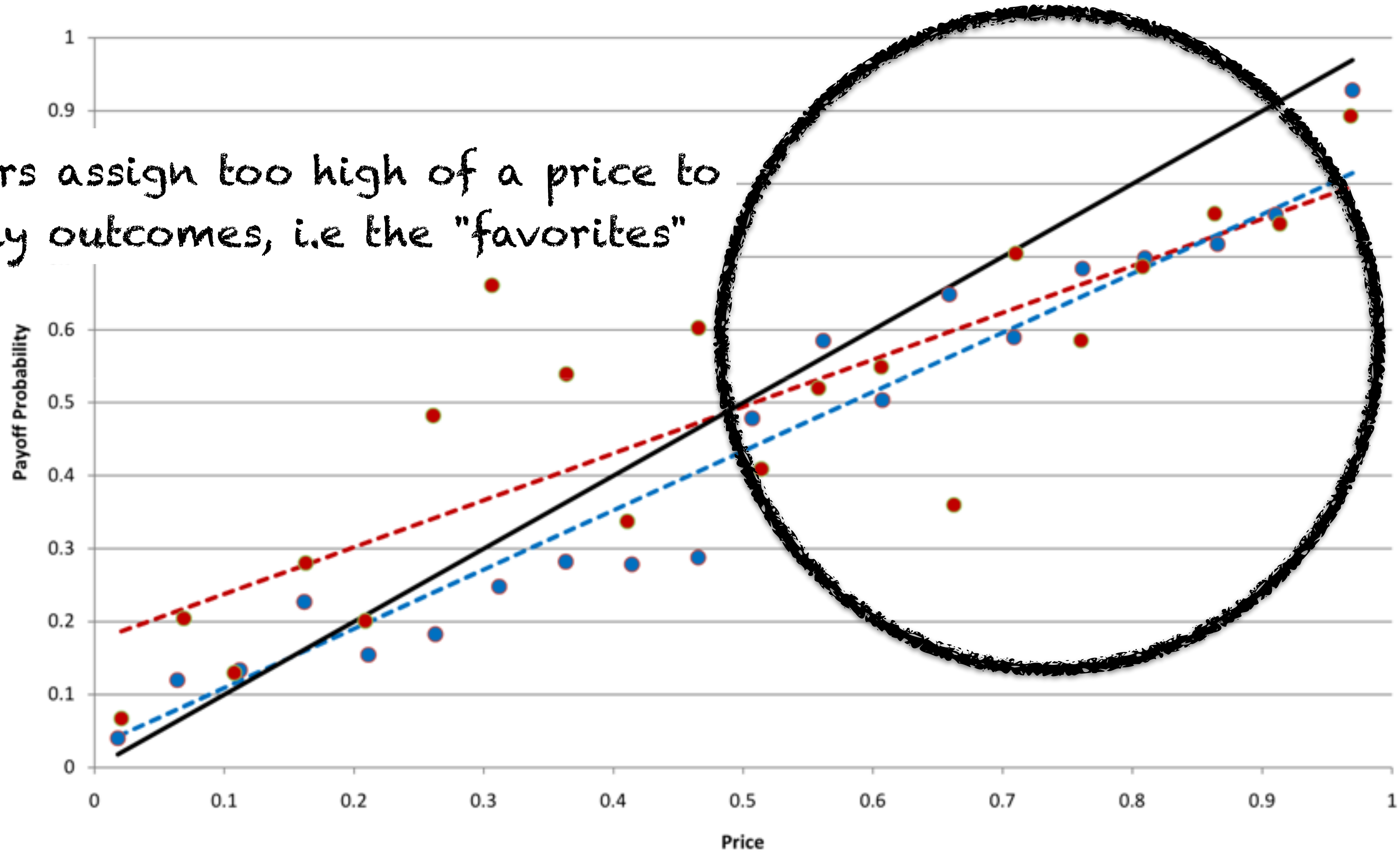
Fig Markets

Traders assign too low of prices to events with low probability. This is a "reverse favorite longshot" bias



Trades in two (red) and five-outcome (blue) markets (22,452and 42,416, respectively) are sorted into 20 bins according to price (i.e., 0-5, 5-10, etc.), and then average price and payoff probability for the bin is plotted. Dashed lines plot regression equations using OLS.

Figure 2. Prices and Probabilities in Two and Five-outcome Markets



Trades in two (red) and five-outcome (blue) markets (22,452and 42,416, respectively) are sorted into 20 bins according to price (i.e., 0-5, 5-10, etc.), and then average price and payoff probability for the bin is plotted. Dashed lines plot regression equations using OLS.

Optimism Biases

- Markets overpriced securities tied to optimistic outcomes by 10 percentage points.
- The optimistic bias was significantly greater on and following days when Google stock appreciated.
- Partly driven by the trading of newly hired employees; employees with longer tenure were better calibrated.

Optimism Biases

- The optimistic bias was largest in:
  - Two outcome markets
  - Early in the sample period
  - Earlier in each quarter.
  - Categories where outcomes are under the control of Google employees i.e. company news (office openings), performance (project completion and product quality).

Table 5. Optimistic bias in the Google markets

	Obs.	Avg price	Avg payoff	Return (SE)	
All markets	70,706	0.357	0.342	-0.015***	(0.003)
Markets with implication for Google	37,910	0.310	0.293	-0.017***	(0.004)
Two-outcome markets with implication for Google	9,023	0.509	0.492	-0.017***	(0.006)
Best outcome for Google	4,556	0.456	0.199	-0.256***	(0.063)
Worst	4,467	0.563	0.790	0.227***	(0.064)
Five-outcome markets with implication for Google	26,511	0.239	0.222	-0.017***	(0.005)
Best outcome for Google	5,592	0.244	0.270	0.027	(0.040)
2nd	5,638	0.271	0.246	-0.025	(0.066)
3rd	5,539	0.296	0.179	-0.118**	(0.053)
4th	5,199	0.206	0.178	-0.028	(0.041)
Worst	4,543	0.162	0.236	0.074	(0.056)

Notes: Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within markets.

Table 5. Optimistic bias in the Google markets

	Obs.	Avg price	Avg payoff	Return (SE)	
All markets	70,706	0.357	0.342	-0.015***	(0.003)
Markets with implication for Google	37,910	0.310	0.293	-0.017***	(0.004)
Two-outcome markets with implication for Google	9,023	0.509	0.492	-0.017***	(0.006)
Best outcome for Google	4,556	0.456	0.199	-0.256***	(0.063)
Worst	4,467	0.563	0.790	0.227***	(0.064)
Five-outcome markets with implication for Google	26,511	0.239	0.222	-0.017***	(0.005)
Best outcome for Google	5,592	0.244	0.270	0.027	(0.040)
2nd	5,638	0.271	0.246	-0.025	(0.066)
3rd	5,539	0.296	0.179	-0.118**	(0.053)
4th	5,199	0.206	0.178	-0.028	(0.041)
Worst	4,543	0.162	0.236	0.074	(0.056)

Notes: Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within markets.

Table 6. Optimism bias by subsample
Dependent variable: returns to expiry
Independent variable: optimism of security (scaled -1 to 1)

Sample	Obs.	Unique markets	Coeff.	S.E.	Constant	S.E.
All markets with implication for Google	37,910	157	-0.105***	(0.036)	-0.013***	(0.004)
Company News	7,430	22	-0.182***	(0.064)	-0.015**	(0.006)
Demand forecasting	12,387	51	-0.042	(0.042)	-0.022***	(0.008)
External News	6,898	42	0.100**	(0.041)	-0.011	(0.009)
Performance (e.g., schedule, product quality)	10,057	38	-0.211***	(0.077)	0.000	(0.010)
2 outcome markets	9,023	50	-0.242	(0.227)	-0.015***	(0.005)
5 outcome markets	26,511	96	-0.013	(0.032)	-0.017***	(0.005)
2005 (Q2 to Q4)	12,224	50	-0.210***	(0.065)	-0.013***	(0.005)
2006 (Q1 to Q4)	20,847	67	-0.026	(0.039)	-0.019***	(0.006)
2007 (Q1 to Q3)	4,839	44	-0.086	(0.066)	-0.007	(0.006)
First month of calendar quarter	15,397	106	-0.121**	(0.054)	-0.010*	(0.006)
Second month	14,234	120	-0.105**	(0.045)	-0.012**	(0.006)
Third month	8,279	105	-0.073**	(0.034)	-0.023**	(0.009)

Notes: Each row is a regression. Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within markets. Optimism is scaled so that the worst outcome for Google is coded -1 and the best is coded 1. I.e., (-1, 1), (-1, 0, 1), (-1, -0.33, 0.33,1), and (-1, -0.5, 0, 0.5, 1) for 2, 3, 4, and 5 outcome markets, respectively.

Table 6. Optimism bias by subsample

Dependent variable: returns to expiry

Independent variable: optimism of security (scaled -1 to 1)

Sample	Obs.	Unique markets	Coeff.	S.E.	Constant	S.E.
All markets with implication for Google	37,910	157	-0.105***	(0.036)	-0.013***	(0.004)
Company News	7,430	22	-0.182***	(0.064)	-0.015**	(0.006)
Demand forecasting	12,387	51	-0.042	(0.042)	-0.022***	(0.008)
External News	6,898	42	0.100**	(0.041)	-0.011	(0.009)
Performance (e.g., schedule, product quality)	10,057	38	-0.211***	(0.077)	0.000	(0.010)
2 outcome markets	9,023	50	-0.242	(0.227)	-0.015***	(0.005)
5 outcome markets	26,511	96	-0.013	(0.032)	-0.017***	(0.005)
2005 (Q2 to Q4)	12,224	50	-0.210***	(0.065)	-0.013***	(0.005)
2006 (Q1 to Q4)	20,847	67	-0.026	(0.039)	-0.019***	(0.006)
2007 (Q1 to Q3)	4,839	44	-0.086	(0.066)	-0.007	(0.006)
First month of calendar quarter	15,397	106	-0.121**	(0.054)	-0.010*	(0.006)
Second month	14,234	120	-0.105**	(0.045)	-0.012**	(0.006)
Third month	8,279	105	-0.073**	(0.034)	-0.023**	(0.009)

Notes: Each row is a regression. Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within markets. Optimism is scaled so that the worst outcome for Google is coded -1 and the best is coded 1. I.e., (-1, 1), (-1, 0, 1), (-1, -0.33, 0.33,1), and (-1, -0.5, 0, 0.5, 1) for 2, 3, 4, and 5 outcome markets, respectively.

Table 9. Regressions predicting trade characteristics from traders' attributes
Dependent variable: Security characteristic*(1 if buy, -1 if sell)

Dependent variable	Optimism (scaled -1 to 1)		Favorite Price - 1/N		Extreme Abs(Optimism)		Buy		Return	
Relationship with returns	Neg.		Neg.		Pos.		Neg.			
Coder? (Participated in code review)	0.033		-0.102	***	-0.284	***	-0.404	***	0.072	***
	(0.049)		(0.022)		(0.081)		(0.139)		(0.023)	
Level (Distance from CEO)	0.006		0.004		0.066	**	0.102	**	0.023	**
	(0.019)		(0.007)		(0.029)		(0.040)		(0.009)	
Hire date (in years)	0.051	**	-0.032	***	-0.093	***	-0.224	***	0.005	
	(0.021)		(0.008)		(0.034)		(0.041)		(0.009)	
NYC-based	-0.169		-0.050	*	0.028		0.014		0.017	
	(0.105)		(0.029)		(0.086)		(0.121)		(0.024)	
Mountain View (MTV)-based	-0.119		-0.101	***	0.161	*	-0.005		0.045	
	(0.105)		(0.031)		(0.096)		(0.122)		(0.029)	
Distance to Noname Café in miles (0 if non-MTV)	0.032		0.085	*	-0.161		-0.597	**	0.069	
	(0.125)		(0.047)		(0.179)		(0.294)		(0.043)	
Experience [Ln(1 + previous trades)]	-0.014		-0.044	***	-0.049	***	-0.094	***	0.026	***
	(0.011)		(0.004)		(0.019)		(0.031)		(0.003)	
Trades	37,910		70,706		37,910		70,706		70,706	
Unique traders	1,126		1,463		1,126		1,463		1,463	

Note: Each observation is a side of a trade. Regressions use trader chatacteristics to predict security characteristics, multipled by -1 if the side in question is a sell. Regressions include trade fixed effects and a dummy variable for one particular extremely prolific trader. Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within person.

New hires more likely to take optimistic positions and more likely to hold short positions, but less likely to over invest in favorites...

Dependent variable	Optimism (scaled -1 to 1)		Favorite Price - 1/N		Extreme Abs(Optimism)		Buy		Return	
Relationship with returns	Neg.		Neg.		Pos.		Neg.			
Coder? (Participated in code review)	0.033		-0.102	***	-0.284	***	-0.404	***	0.072	***
	(0.049)		(0.022)		(0.081)		(0.139)		(0.023)	
Level (Distance from CEO)	0.006		0.004		0.066	**	0.102	**	0.023	**
	(0.019)		(0.007)		(0.029)		(0.040)		(0.009)	
Hire date (in years)	0.051	**	-0.032	***	-0.093	***	-0.224	***	0.005	
	(0.021)		(0.008)		(0.034)		(0.041)		(0.009)	
NYC-based	-0.169		-0.050	*	0.028		0.014		0.017	
	(0.105)		(0.029)		(0.086)		(0.121)		(0.024)	
Mountain View (MTV)-based	-0.119		-0.101	***	0.161	*	-0.005		0.045	
	(0.105)		(0.031)		(0.096)		(0.122)		(0.029)	
Distance to Noname Café in miles (0 if non-MTV)	0.032		0.085	*	-0.161		-0.597	**	0.069	
	(0.125)		(0.047)		(0.179)		(0.294)		(0.043)	
Experience [Ln(1 + previous trades)]	-0.014		-0.044	***	-0.049	***	-0.094	***	0.026	***
	(0.011)		(0.004)		(0.019)		(0.031)		(0.003)	
Trades	37,910		70,706		37,910		70,706		70,706	
Unique traders	1,126		1,463		1,126		1,463		1,463	

Note: Each observation is a side of a trade. Regressions use trader chatacteristics to predict security characteristics, multipled by -1 if the side in question is a sell. Regressions include trade fixed effects and a dummy variable for one particular extremely prolific trader. Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within person.

More experienced traders are more likely to trade against the market's biases...

Dependent variable	Optimism (scaled -1 to 1)		Favorite Price - 1/N		Extreme Abs(Optimism)		Buy		Return	
Relationship with returns	Neg.		Neg.		Pos.		Neg.			
Coder? (Participated in code review)	0.033		-0.102	***	-0.284	***	-0.404	***	0.072	***
	(0.049)		(0.022)		(0.081)		(0.139)		(0.023)	
Level (Distance from CEO)	0.006		0.004		0.066	**	0.102	**	0.023	**
	(0.019)		(0.007)		(0.029)		(0.040)		(0.009)	
Hire date (in years)	0.051	**	-0.032	***	-0.093	***	-0.224	***	0.005	
	(0.021)		(0.008)		(0.034)		(0.041)		(0.009)	
NYC-based	-0.169		-0.050	*	0.028		0.014		0.017	
	(0.105)		(0.029)		(0.086)		(0.121)		(0.024)	
Mountain View (MTV)-based	-0.119		-0.101	***	0.161	*	-0.005		0.045	
	(0.105)		(0.031)		(0.096)		(0.122)		(0.029)	
Distance to Noname Café in miles (0 if non-MTV)	0.032		0.085	*	-0.161		-0.597	**	0.069	
	(0.125)		(0.047)		(0.179)		(0.294)		(0.043)	
Experience [Ln(1 + previous trades)]	-0.014		-0.044	***	-0.049	***	-0.094	***	0.026	***
	(0.011)		(0.004)		(0.019)		(0.031)		(0.003)	
Trades	37,910		70,706		37,910		70,706		70,706	
Unique traders	1,126		1,463		1,126		1,463		1,463	

Note: Each observation is a side of a trade. Regressions use trader chatacteristics to predict security characteristics, multipled by -1 if the side in question is a sell. Regressions include trade fixed effects and a dummy variable for one particular extremely prolific trader. Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within person.

Correlations

- Study information flows using measures of proximity: Geographical, Organizational, Social, Demographic
- Predict the size and direction of the trades from the prior positions of proximate colleagues

Table 10. Geography and trading correlations

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)		(2)		(3)		(4)		(5)		(6)	
Geographical proximity												
Same city	0.006		0.000		0.003		-0.001		-0.002		-0.002	
	(0.004)		(0.006)		(0.007)		(0.006)		(0.006)		(0.006)	
Proximity within city			0.010	*	-0.004		-0.014	*	-0.014	*	-0.013	
(100ft/distance between buildings, min = 0, max = 1)			(0.006)		(0.007)		(0.008)		(0.008)		(0.008)	
Same building					0.022	***	0.008		-0.001		0.002	
					(0.005)		(0.007)		(0.007)		(0.007)	
Same floor							0.025	***	-0.019	*	-0.020	*
							(0.009)		(0.010)		(0.010)	
Proximity on floor									0.090	***	0.053	***
(10ft/distance between offices, min = 0, max = 1)									(0.015)		(0.017)	
Same office											0.055	***
											(0.016)	
Building information missing for either party			-0.004		-0.005		0.000		0.000		-0.001	
			(0.005)		(0.005)		(0.006)		(0.006)		(0.005)	
Room information missing for either party							-0.021	***	-0.025	***	-0.025	***
							(0.008)		(0.008)		(0.008)	
Other controls												
Trade fixed effects	X		X		X		X		X		X	
Initial position	X		X		X		X		X		X	
Observations	140,768		140,768		140,768		140,768		140,768		140,768	
R-squared	0.0352		0.0354		0.0359		0.0378		0.0395		0.0399	

Notes: Independent variables are the sum of the pre-trade position of a trader's colleagues, weighted by the variable given (e.g., an indicator for whether the two traders are in the same city). Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within person.

Table 10. Geography and trading correlations

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)	(2)	(3)	(4)	(5)	(6)
Geographical proximity						
Same city	0.006	0.000	0.003	-0.001	-0.002	-0.002
	(0.004)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
Proximity within city		0.010 *	-0.004	-0.014 *	-0.014 *	-0.013
(100ft/distance between buildings, min = 0, max = 1)		(0.006)	(0.007)	(0.008)	(0.008)	(0.008)
Same building			0.022 ***	0.008	-0.001	0.002
			(0.005)	(0.007)	(0.007)	(0.007)
Same floor						
Proximity on floor					[illegible]	[illegible]
(10ft/distance between offices, min = 0, max = 1)					(0.015)	(0.017)
Same office						0.055 ***
						(0.016)
Building information missing for either party		-0.004	-0.005	0.000	0.000	-0.001
		(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Room information missing for either party				-0.021 ***	-0.025 ***	-0.025 ***
				(0.008)	(0.008)	(0.008)
Other controls						
Trade fixed effects	X	X	X	X	X	X
Initial position	X	X	X	X	X	X
Observations	140,768	140,768	140,768	140,768	140,768	140,768
R-squared	0.0352	0.0354	0.0359	0.0378	0.0395	0.0399

Most correlation between employees sharing an office

Notes: Independent variables are the sum of the pre-trade position of a trader's colleagues, weighted by the variable given (e.g., an indicator for whether the two traders are in the same city). Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within person.

Table 10. Geography and trading correlations

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)	(2)	(3)	(4)	(5)	(6)
Geographical proximity						
Same city	0.006	0.000	0.003	-0.001	-0.002	-0.002
	(0.004)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
Proximity within city		0.010 *	-0.004	-0.014 *	-0.014 *	-0.013
(100ft/distance between buildings, min = 0, max = 1)		(0.006)	(0.007)	(0.008)	(0.008)	(0.008)
						*
				(0.009)	(0.010)	(0.010)
Proximity on floor					0.090 ***	0.053 ***
(10ft/distance between offices, min = 0, max = 1)					(0.015)	(0.017)
Same office						0.055 ***
						(0.016)
Building information missing for either party		-0.004	-0.005	0.000	0.000	-0.001
		(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Room information missing for either party				-0.021 ***	-0.025 ***	-0.025 ***
				(0.008)	(0.008)	(0.008)
Other controls						
Trade fixed effects	X	X	X	X	X	X
Initial position	X	X	X	X	X	X
Observations	140,768	140,768	140,768	140,768	140,768	140,768
R-squared	0.0352	0.0354	0.0359	0.0378	0.0395	0.0399

Correlation decreases with distance, even on the same floor

Notes: Independent variables are the sum of the pre-trade position of a trader's colleagues, weighted by the variable given (e.g., an indicator for whether the two traders are in the same city). Standard errors are heteroskedasticity robust and adjust for clustering of outcomes within person.

Table 11. Social and work relationships and correlated trading

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)	(2)	(3)	(4)	(5)	(6)
Social connections						
Self-reported professional relationship?	0.016 *	0.009	0.010	0.012	0.017	0.020 *
	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(0.011)
Self-reported friendship?	-0.001	-0.044 **	-0.050 **	-0.050 **	-0.040 *	-0.054 **
	(0.019)	(0.021)	(0.020)	(0.021)	(0.022)	(0.023)
# of overlapping email lists	0.000	-0.001	-0.003	-0.004	-0.005	-0.007
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Work history						
Reviewed each other's code		0.028 ***	0.027 ***	0.019 **	0.023 **	0.017 *
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Overlapped on project?		0.034 ***	0.010	-0.031 **	-0.050 ***	-0.026
		(0.012)	(0.014)	(0.015)	(0.016)	(0.016)
Organizational proximity						
Same SVP (one level below CEO)			0.016 ***	0.014 **	0.015 ***	0.015 **
			(0.006)	(0.006)	(0.006)	(0.006)
Same "2-Levels-below-CEO" manager			-0.011 *	-0.008	-0.007	-0.007
			(0.006)	(0.008)	(0.008)	(0.008)
Same "3-Levels-below-CEO" manager			0.033 **	-0.018	-0.026	-0.026
			(0.014)	(0.017)	(0.017)	(0.017)
1-2 steps away on org chart				0.102 ***	0.061 ***	0.068 ***
				(0.018)	(0.017)	(0.017)
3 steps away on org chart				-0.016	-0.020 *	-0.019 *
				(0.011)	(0.011)	(0.011)
Other controls						
Trade fixed effects	X	X	X	X	X	X
Initial position	X	X	X	X	X	X
Geographic proximity variables (from Table 10, cols 6)					X	X
Demographic similarity						X
Observations	140,768	140,768	140,768	140,768	140,768	140,768
R-squared	0.035	0.0357	0.0372	0.0392	0.0423	0.0433

Notes: The last column includes 8 variables capturing the pre-trade positions of colleagues who are similar along a demographic dimension (attended the same undergraduate school, had the same undergrad major, are both or neither in programming roles at Google, are both native English speakers, share a common non-English native language, or are similar according to three commonly studied demographic variables).

Table 11. Social and work relationships and correlated trading

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)	(2)	(3)	(4)	(5)	(6)
Social connections						
Self-reported professional relationship?	0.016 *	0.009	0.010	0.012	0.017	0.020 *
	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(0.011)
Self-reported friendship?	-0.001	-0.044 **	-0.050 **	-0.050 **	-0.040 *	-0.054 **
	(0.019)	(0.021)	(0.020)	(0.021)	(0.022)	(0.023)
# of overlapping email lists	0.000	-0.001	-0.003	-0.004	-0.005	-0.007
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Work history						
Reviewed each other's code		0.028 ***	0.027 ***	0.019 **	0.023 **	0.017 *
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Overlapped on project?		0.034 ***	0.010	-0.031 **	-0.050 ***	-0.026
		(0.012)	(0.014)	(0.015)	(0.016)	(0.016)
Organizational proximity						
Trade fixed effects	X	X	X	X	X	X
Initial position	X	X	X	X	X	X
Geographic proximity variables (from Table 10, cols 6)					X	X
Demographic similarity						X
Observations	140,768	140,768	140,768	140,768	140,768	140,768
R-squared	0.035	0.0357	0.0372	0.0392	0.0423	0.0433

"We find that measures of social connections, either self-reported on the April 2006 survey or inferred from subscriptions to email lists, do not explain trading correlations well. A history of reviewing each other's code or overlapping on a project does, however."

Notes: The last column includes 8 variables capturing the pre-trade positions of colleagues who are similar along a demographic dimension (attended the same undergraduate school, had the same undergrad major, are both or neither in programming roles at Google, are both native English speakers, share a common non-English native language, or are similar according to three commonly studied demographic variables).

"The single best explanator is being within one or two steps on the organization chart (i.e., sharing a manager, being someone's manager, or being someone's manager's manager)."

Table 11. Social and work relationships and correlated trading

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)	(2)	(3)	(4)	(5)	(6)
Social connections						
Self-reporte[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.020 *
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	(0.011)
Self-reporte[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.054 **
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	(0.023)
# of overlap[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.007
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	(0.005)
Work history						
Reviewed e[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.017 *
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	(0.009)
Overlapped on project?		[illegible]	[illegible]	[illegible]	[illegible]	-0.026
		(0.012)	(0.014)	(0.015)	(0.016)	(0.016)
Organizational proximity						
Same SVP (one level below CEO)			0.016 ***	0.014 **	0.015 ***	0.015 **
			(0.006)	(0.006)	(0.006)	(0.006)
Same "2-Levels-below-CEO" manager			-0.011 *	-0.008	-0.007	-0.007
			(0.006)	(0.008)	(0.008)	(0.008)
Same "3-Levels-below-CEO" manager			0.033 **	-0.018	-0.026	-0.026
			(0.014)	(0.017)	(0.017)	(0.017)
1-2 steps away on org chart				0.102 ***	0.061 ***	0.068 ***
				(0.018)	(0.017)	(0.017)
3 steps away on org chart				-0.016	-0.020 *	-0.019 *
				(0.011)	(0.011)	(0.011)
Other controls						
Trade fixed effects	X	X	X	X	X	X
Initial position	X	X	X	X	X	X
Geographic proximity variables (from Table 10, cols 6)					X	X
Demographic similarity						X
Observations	140,768	140,768	140,768	140,768	140,768	140,768
R-squared	0.035	0.0357	0.0372	0.0392	0.0423	0.0433

Notes: The last column includes 8 variables capturing the pre-trade positions of colleagues who are similar along a demographic dimension (attended the same undergraduate school, had the same undergrad major, are both or neither in programming roles at Google, are both native English speakers, share a common non-English native language, or are similar according to three commonly studied demographic variables).

Table 11. Social and work relationships and correlated trading

Dependent variable: net shares purchased (normalized)

Independent variables: Proximity-weighted normalized sums of colleagues' pre-trade positions

	(1)	(2)	(3)	(4)	(5)	(6)
Social connections						
Self-reported professional relationship?	0.016 *	0.009	0.010	0.012	0.017	0.020 *
	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(0.011)
Self-reported friendship?	-0.001	-0.044 **	-0.050 **	-0.050 **	-0.040 *	-0.054 **
	(0.019)	(0.021)	(0.020)	(0.021)	(0.022)	(0.023)
# of overlapping email lists	0.000	-0.001	-0.003	-0.004	-0.005	-0.007
						(0.005)
Work history						
Reviewed ea					*	0.017 *
						(0.009)
Overlapped					**	-0.026
						(0.016)
Organizational pr						
Same SVP (o					**	0.015 **
						(0.006)
Same "2-Lev						-0.007
						(0.008)
Same "3-Lev						-0.026
						(0.017)
1-2 steps aw					**	0.068 ***
						(0.017)
3 steps away						-0.019 *
						(0.011)
Other controls						
Trade fixed e						X
Initial positi						X
Geographic						X
Demographi						X
Observations	140,768	140,768	140,768	140,768	140,768	140,768
R-squared	0.035	0.0357	0.0372	0.0392	0.0423	0.0433

"...employees most likely to have correlated trading are those who are proximate organizationally or geographically and are not friends. One admittedly speculative interpretation of this result is that friends have better things to discuss than the subjects of prediction markets, while the prediction markets provide a topic of conversation for those who are not friends."

Notes: The last column includes 8 variables capturing the pre-trade positions of colleagues who are similar along a demographic dimension (attended the same undergraduate school, had the same undergrad major, are both or neither in programming roles at Google, are both native English speakers, share a common non-English native language, or are similar according to three commonly studied demographic variables).

Prediction Market Take Aways

- Prediction markets are simple securities markets that allow traders to profit from correct private information about the outcomes of future events
- Individuals' desires to make money allows the market to aggregate all of the traders' beliefs, reflected in the price
- These markets have been shown to behave efficiently, and provide correct predictions with high accuracy
- Markets can be used by companies and researchers to make business decisions, study organizational structures, and measure social networks
- Using prediction markets for this kind of research is more "real-time" and possibly more accurate than using retrospective surveys

Good bye, all.
Go forth and do.

Sources

Prediction Markets (Justin Wolfers and Eric Zitzewitz)
http://www.nber.org/papers/w10504.pdf

Using Prediction Markets to Track Information Flows: Evidence from Google (Bo Cowgill, Justin Wolfers, and Eric Zitzewitz)
http://www.eecs.harvard.edu/cs286r/courses/fall10/papers/GooglePredictionMarketPaper.pdf