Class 15: Experimental studies of contagion

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University

Last class:

- differentiate between simple and complex contagions (require 2 or more neighbors to be active)

Last class:

- differentiate between simple and complex contagions (require 2 or more neighbors to be active)
- changes in network structure that promote simple contagion don't always promote complex contagion

Last class:

- differentiate between simple and complex contagions (require 2 or more neighbors to be active)
- changes in network structure that promote simple contagion don't always promote complex contagion
- experiments showing that some behaviors spread more and faster in highly clustered networks rather than random networks

Complementary approach to contagion:

1. Nickerson, D.W. (2008). Is voting contagious? Evidence from two field experiments. *American Political Science Review*.
2. Kramer, A.D.I. et al. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*.

Community minute

3 pieces of general context:

- "birds of a feather flock together" but why?
- Experiments are powerful ways to isolate and estimate causal effects
- Experiments are powerful but not perfect: internal validity, external validity, and ethics

3 pieces of general context:

- "birds of a feather flock together" but why?
- Experiments are powerful ways to isolate and estimate causal effects
- Experiments are powerful but not perfect: internal validity, external validity, and ethics

People are often similar to the people they are connected to socially

People are often similar to the people they are connected to socially

- selection (like people become friends)

People are often similar to the people they are connected to socially

- selection (like people become friends)
- shared environment

People are often similar to the people they are connected to socially

- selection (like people become friends)
- shared environment
- contagion

People are often similar to the people they are connected to socially

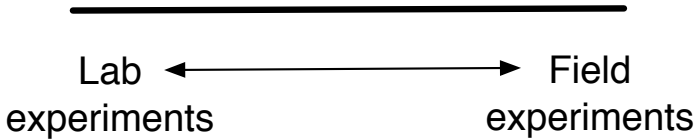
- selection (like people become friends)
- shared environment
- contagion

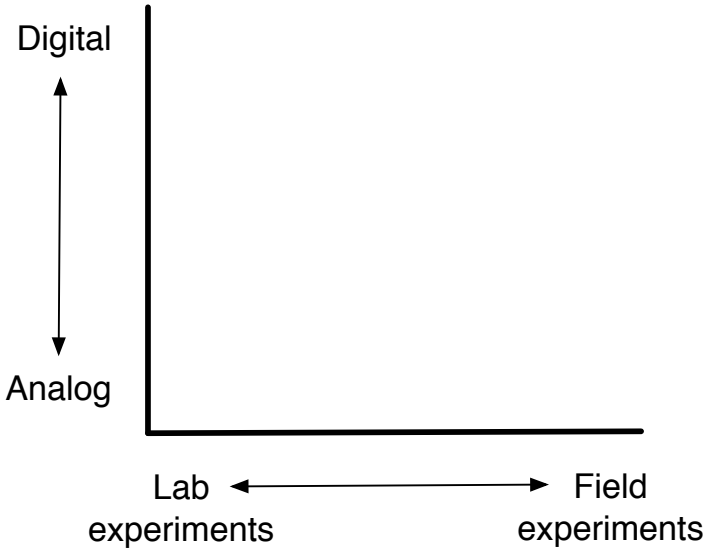
For some traits, one factor might dominate, but for many traits all might be at work.

- "birds of a feather flock together" but why?
- Experiments are powerful ways to isolate and estimate causal effects
- Experiments are powerful but not perfect: internal validity, external validity, and ethics

"It's like you don't harass women, you don't steal, and you've got to have a control group. This is one of the things that you can lose your job for at Harrah's not running a control group." Gary Loveman, CEO Harrah's

It is hard to make causal claims without an experiment, as both papers describe. Part of the contribution of each paper is to bring experimental evidence. Here we saw two *field* experiments.


Lab
experiments
Field
experiments


Digital
Analog
Lab
experiments
Field
experiments

Experiments have four main ingredients:

- recruiting participants

Experiments have four main ingredients:

- recruiting participants
- randomization treatment

Experiments have four main ingredients:

- recruiting participants
- randomization treatment
- delivering treatment and control

Experiments have four main ingredients:

- recruiting participants
- randomization treatment
- delivering treatment and control
- measuring outcomes

Both experiments you read put a lot of care into creating the control group.

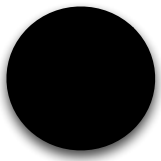
- Voting paper by Nickerson: compared Get-out-the-Vote message to Recycling message

Both experiments you read put a lot of care into creating the control group.

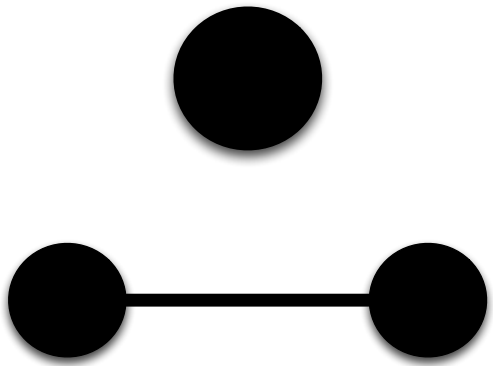
- Voting paper by Nickerson: compared Get-out-the-Vote message to Recycling message
- Emotional contagion paper by Kramer et al.: a control group for positivity reduced condition and a control group for negatively reduced condition because of different base rates (e.g., 22.4% of posted had negative words, 46.8% had positive words)

A note on terminology:
Perturb and observe experiments vs randomized controlled experiments

These experiments move from the individual to the dyad.



These experiments move from the individual to the dyad.



- "birds of a feather flock together" but why?
- Experiments are powerful ways to isolate and estimate causal effects
- Experiments are powerful but not perfect: internal validity, external validity

You saw internal validity and external validity were categories to organize concerns. Explicit in Nickerson study on voting; implicit in Kramer et al. study of emotions

- "birds of a feather flock together" but why?
- Experiments are powerful ways to isolate and estimate causal effects
- Experiments are powerful but not perfect: internal validity, external validity

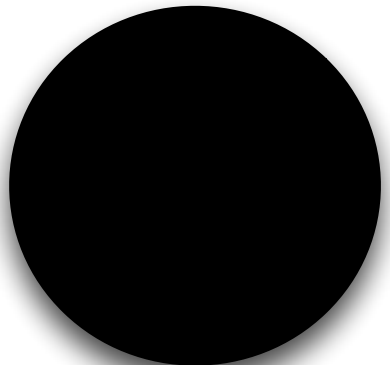
Given that common background let's dive in

American Political Science Review Vol. 102, No. 1 February 2008
DOI: 10.1017/S0003055408080039

Is Voting Contagious? Evidence from Two Field Experiments

DAVID W. NICKERSON *University of Notre Dame*

http://doi.org/10.1017/S0003055408080039



$$P_1 \Leftrightarrow P_2$$

$$\begin{array}{c} \textit{Treatment} \\ \downarrow^{T} \\ P_1 \overset{S}{\rightarrow} P_2 \end{array}$$

Contagion effect: $\alpha = \frac{S}{T}$

Note that there is nothing specific in this design to voting. This could be any intervention.

TABLE 1. Possible Outcomes under placebo protocol

		Probability of Event Occurring	Voting Rate of Answerer	Voting Rate of Person Who Did Not Answer Door
GOTV	Door Answered	π	$\mu_1 + T$	$\mu_2 + S$
	No Answer	$1 - \pi$	N.A.[a]	μ_3
Recycling	Door Answered	π	μ_1	μ_2
	No Answer	$1 - \pi$	N.A.	μ_3

[a] N.A. = Not applicable.

The role of the recycling intervention is to create a "fair" comparison.

TABLE 3. Treatment Effect among Contacted Households

	Denver		Minneapolis		Pooled	
	Direct	Secondary	Direct	Secondary	Direct	Secondary
Percent Voting in GOTV Group	47.7% (3.0)	42.4% (2.9)	27.1% (3.1)	23.6% (3.0)		
Percent Voting in Recycling Group	39.1% (2.9)	36.9% (2.9)	16.2% (2.7)	17.3% (2.7)		
Estimated Treatment Effect	**8.6% (4.2)**	**5.5% (4.1)**	**10.9% (4.1)**	**6.4% (4.1)**	**9.8% (2.9)**	**6.0% (2.9)**
P-Value	0.02	0.09	<0.01	0.06	<0.01	0.02

Note. Numbers in parentheses represent standard errors. P-values test the one-tailed hypothesis. Pooled estimates are weighted averages of results for both cities.

How much should you trust these results? Internal and external validity

TABLE 2. Balance of Observable Traits by Treatment Assignment

		Denver			Minneapolis		
Stage	Category	GOTV	Recycling	Control	GOTV	Recycling	Control
Assignment	Age	56.1	55.5	56.1	46.6	47.9	45.9
	Votes cast in past five elections	2.9	2.8	2.9	2.6	2.6	2.6
Application	House Contacted	33.2%	32.8%		46.2%	43.5%	
	Go Away	2.5%	4.1%		1.8%	1.1%	
	Moved	0.9%	0.6%		1.4%	0.7%	
	Can't Attempt	5.4%	4.2%		6.6%	6.4%	
	No Answer	58.0%	58.3%		44.0%	48.3%	
	Number Contacted	283	279		203	191	
Contacted	Age	55.9	56.0		47.7	48.5	
	Votes cast in past five elections	2.9	2.9		2.7	2.7	

Note. Age and vote history were taken from county voter files. Canvassers were asked to record the disposition of each door knock.

Internal validity: it looks like the get-out-the-vote people and the recycling people are similar

External validity (partial list)

- ties within households are different from other ties

External validity (partial list)

- ties within households are different from other ties
- households in this study might be different from other households

External validity (partial list)

- ties within households are different from other ties
- households in this study might be different from other households
- these results are from a low salience election (might be different in a presidential election)

External validity (partial list)

- ties within households are different from other ties
- households in this study might be different from other households
- these results are from a low salience election (might be different in a presidential election)
- other behaviors might not be as contagious as voter turnout

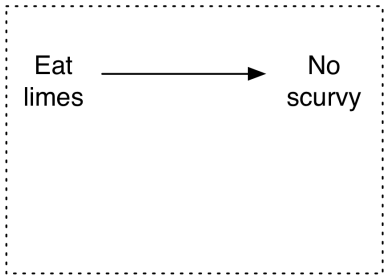
Notes on application:

Notes on application:

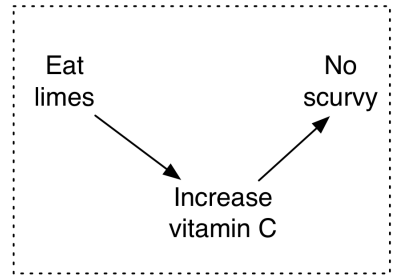
- Need to count the spillover (if you generate 100 direct votes, you also generate about 60 indirect votes)

Notes on application:

- Need to count the spillover (if you generate 100 direct votes, you also generate about 60 indirect votes)
- No idea about mechanism so hard to design more contagious treatments



Causal effect without mechanism



Causal effect with mechanism

- contagion of voting via "intervene and spillover" design

- contagion of voting via "intervene and spillover" design
- contagion of emotion via an "edge-control" design


SAVEMONEY

How does the content you see on social media impact your emotions?

https://unsplash.com/search/photos/party?photo=uD0W-swVGgE

Simplify: How does seeing happy content from your friend impact you?

Simplify: How does seeing happy content from your friend impact you?

- Seeing your friends doing happy things will make you happy (contagion)

Simplify: How does seeing happy content from your friend impact you?

- Seeing your friends doing happy things will make you happy (contagion)
- Seeing your friends doing happy things will make you sad (relative deprivation)

Non-experimental approach:

Non-experimental approach:

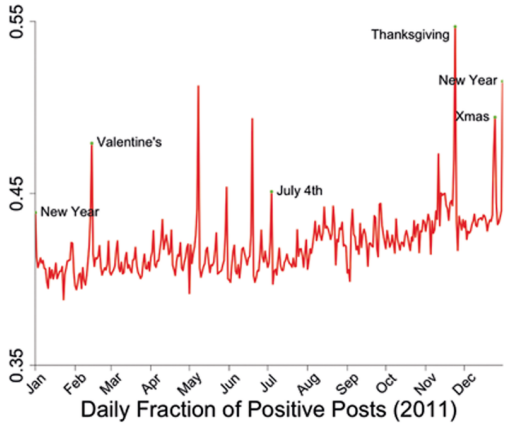
- Classify all posts as happy or sad based on the words that they use

Non-experimental approach:

- Classify all posts as happy or sad based on the words that they use
- Count the proportion of your posts that are positive after your friends make positive posts

Non-Experimental approach falls victim to the Thanksgiving trap!

Non-Experimental approach falls victim to the Thanksgiving trap!



Coviello, et al 2014, Fig 1A

Possible solution: Experiment

Possible solution: Experiment

How could you possibly precisely control the emotional content to which people are exposed and then measure the outcomes?

Possible solution: Experiment

How could you possibly precisely control the emotional content to which people are exposed and then measure the outcomes?

You could work at Facebook.

https://www.wired.com/2015/03/facebook-moves-new-garden-roofed-fantasyland/

Experimental approach

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of [b]Communication and [c]Information Science, Cornell University, Ithaca, NY 14853

http://doi.org/10.1073/pnas.1320040111

This design works by changing the edge, not by intervene and spillover

Experimental design:

- 700,000 people

Experimental design:

- 700,000 people
- four arms: positive posts reduced vs control, negative posts reduced vs control

Experimental design:

- 700,000 people
- four arms: positive posts reduced vs control, negative posts reduced vs control
- incoming post scored as positive or negative if they had one positive or negative word as defined by Linguistic Inventory and Word Count (LIWC) (e.g., good, sad, etc)

Experimental design:

- 700,000 people
- four arms: positive posts reduced vs control, negative posts reduced vs control
- incoming post scored as positive or negative if they had one positive or negative word as defined by Linguistic Inventory and Word Count (LIWC) (e.g., good, sad, etc)
- posts randomly blocked from NewsFeed depending on condition (blocking not boosting)

Experimental design:

- 700,000 people
- four arms: positive posts reduced vs control, negative posts reduced vs control
- incoming post scored as positive or negative if they had one positive or negative word as defined by Linguistic Inventory and Word Count (LIWC) (e.g., good, sad, etc)
- posts randomly blocked from NewsFeed depending on condition (blocking not boosting)
- outcome: percentage of words posted that were positive or negative

This exact design requires cooperation from Facebook. More generally, many studies of social media's impact require cooperation from social media companies.

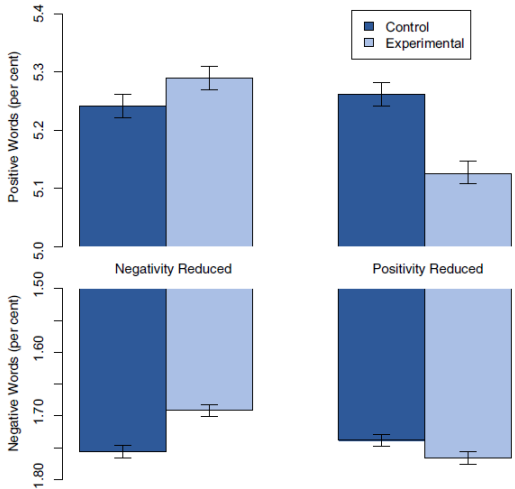
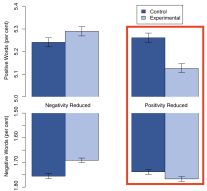
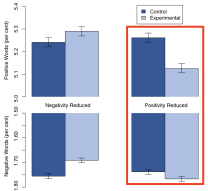


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.



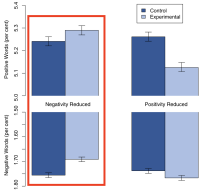
- % positive words in positivity reduced treatment: $\sim$ 5.13% (5.13 words per 100)
- % positive words in positivity reduced control: $\sim$5.27% (5.27 words per 100)
- Difference % positive words: $\sim$-0.14% (0.14 words per 100, 14 words per 10,000)



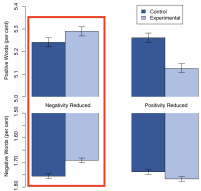
- % positive words in positivity reduced treatment: $\sim$ 5.13% (5.13 words per 100)
- % positive words in positivity reduced control: $\sim$5.27% (5.27 words per 100)
- Difference % positive words: $\sim$-0.14% (0.14 words per 100, 14 words per 10,000)

- % negative words in positivity reduced treatment: $\sim$1.76% (1.76 words per 100)
- % negative words in positivity reduced control: $\sim$1.74% (1.74 words per 100)
- Difference % negative words: $\sim$0.02% (0.02 words per 100, 2 words per 10,000)

Note: These are approximate because I just read them off the graph. In the paper they report the results of a more complex analysis



- % positive words in negativity reduced treatment: ∼5.29% (5.29 words per 100)
- % of positive words in negativity reduced control: ∼5.23% (5.23 words per 100)
- Difference % positive words: ∼0.06% (0.06 words per 100, 6 words per 10,000)



- % positive words in negativity reduced treatment: ∼5.29% (5.29 words per 100)
- % of positive words in negativity reduced control: ∼5.23% (5.23 words per 100)
- Difference % positive words: ∼0.06% (0.06 words per 100, 6 words per 10,000)

- % negative words in negativity reduced treatment: ∼1.69% (1.69 words per 100)
- % negative words in negativity reduced control: ∼1.76% (1.76 words per 100)
- Difference % negative words: ∼-0.07% (0.07 words per 100, 7 words per 10,000)

Note: These are approximate because I just read them off the graph. In the paper they report the results of a more complex analysis

How much should you trust these results? Internal and external validity

Internal validity:

- Was the randomization delivered correctly?
- Was the outcome measured correctly on the right people?

Internal validity:

- Was the randomization delivered correctly?
- Was the outcome measured correctly on the right people?

Designing and Deploying Online Field Experiments

Eytan Bakshy
Facebook
Menlo Park, CA
eytan@fb.com

Dean Eckles
Facebook
Menlo Park, CA
deaneckles@fb.com

Michael S. Bernstein
Stanford University
Palo Alto, CA
msb@cs.stanford.edu

https://arxiv.org/pdf/1409.3174v1.pdf

External validity

External validity:

- Are Facebook posts a good measure of how we feel?

External validity

External validity:

- Are Facebook posts a good measure of how we feel?
- Is word counts a good way to quantify the emotional content of posts? ("I am so so happy" vs "I wish I was happy")

External validity

External validity:

- Are Facebook posts a good measure of how we feel?
- Is word counts a good way to quantify the emotional content of posts? ("I am so so happy" vs "I wish I was happy")

Probably a bad measure of a bad signal

Three other important things about this experiment

- unintended impact of treatment

People who had positivity reduced and people who had negativity reduced, posted fewer words.

People who had positivity reduced and people who had negativity reduced, posted fewer words.

Your treatment can effect your outcome, but also many other outcomes

We also observed a withdrawal effect: People who were exposed to fewer emotional posts (of either valence) in their News Feed were less expressive overall on the following days, addressing the question about how emotional expression affects social engagement online. This observation, and the fact that

Imagine that you work at Facebook and your metric was to increase engagement.

We also observed a withdrawal effect: People who were exposed to fewer emotional posts (of either valence) in their News Feed were less expressive overall on the following days, addressing the question about how emotional expression affects social engagement online. This observation, and the fact that

Imagine that you work at Facebook and your metric was to increase engagement. Would you adjust the NewsFeed to show more emotional content, either accidentally or intentionally?

Three other important things about this experiment

- unintended impact of treatment
- “significance”

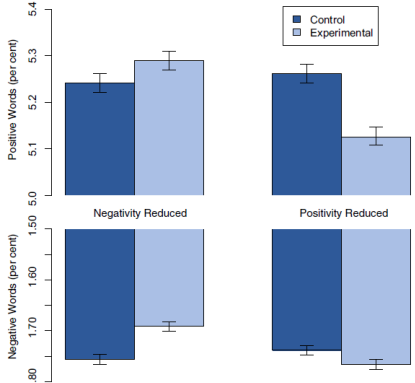


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

- Are differences that size possible due to chance? (statistical significance)

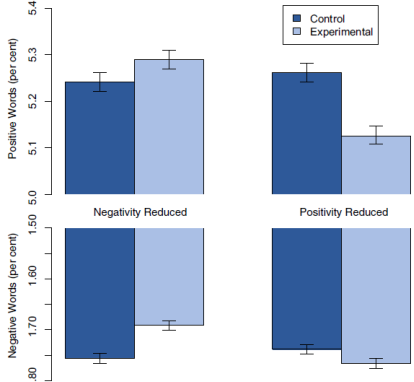


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

- Are differences that size possible due to chance? (statistical significance)
- Are differences that big important? (practical importance)

Although these data provide, to our knowledge, some of the first experimental evidence to support the controversial claims that emotions can spread throughout a network, the effect sizes from the manipulations are small (as small as $d = 0.001$). These effects nonetheless matter given that the manipulation of the independent variable (presence of emotion in the News Feed) was minimal whereas the dependent variable (people's emotional expressions) is difficult to influence given the range of daily experiences that influence mood (10). More importantly, given the massive scale of social networks such as Facebook, even small effects can have large aggregated consequences (14, 15): For example, the well-documented connection between emotions and physical well-being suggests the importance of these findings for public health. Online messages influence our experience of emotions, which may affect a variety of offline behaviors. And after all, an effect size of $d = 0.001$ at Facebook's scale is not negligible: In early 2013, this would have corresponded to hundreds of thousands of emotion expressions in status updates per day.

Three other important things about this experiment

- unintended impact of treatment
- "significance"
- ethics of running this kind of experiment

The Opinion Pages | OP-ED CONTRIBUTOR

Should Facebook Manipulate Users?

Jaron Lanier on Lack of Transparency in Facebook Study

By JARON LANIER JUNE 30, 2014

https://www.nytimes.com/2014/07/01/opinion/jaron-lanier-on-lack-of-transparency-in-facebook-study.html

Stop complaining about the Facebook study. It's a golden age for research

Duncan J Watts

We should *insist* that Facebook do experiments on the decisions it's already making for us. Anything else would be unethical

https://www.theguardian.com/commentisfree/2014/jul/07/facebook-study-science-experiment-research

Chapter 6, Ethics: `http://www.bitbybitbook.com/en/ethics/`

Experiments are powerful but they are not perfect

Experiments are powerful but they are not perfect

- Powerful: enable us to estimate causal effects (avoid Thanksgiving trap)

Experiments are powerful but they are not perfect

- Powerful: enable us to estimate causal effects (avoid Thanksgiving trap)
- Not perfect:
  - potential problems with internal validity
  - potential problems with external validity
  - potential problems with ethics

Summary:

- experimental approaches can measure the effect we have on each other

Summary:

- experimental approaches can measure the effect we have on each other
- voting is contagious & emotional valence of word use is contagious

Summary:

- experimental approaches can measure the effect we have on each other
- voting is contagious & emotional valence of word use is contagious
- two designs: 1) intervene and spillover; 2) edge-control

Summary:

- experimental approaches can measure the effect we have on each other
- voting is contagious & emotional valence of word use is contagious
- two designs: 1) intervene and spillover; 2) edge-control
- some of these experiments raise ethical questions (e.g., Kramer et al.)

Going viral

Have you ever said that something was going viral?

- Goel, S. et al. (2016). The structural virality of online diffusion. *Management Science*
- Cheng et al. (2014) Can cascades be predicted? *WWW*