

Don't @ Me: Experimentally Reducing Partisan Incivility on Twitter

Kevin Munger

NYU

April 21, 2017

Prepared for King's College, April 2017

Cross-Cutting Social Identities



Veteran [redacted] @ [redacted] · 11h

@realDonaldTrump a nasty woman beat the piss out of you in front of millions, yet u claim victory. You are a bad hombre



62



437



1st Sgt. [redacted] @ [redacted] · 11h

@ [redacted] @realDonaldTrump repeat that statement when ure head gets sawed off.



3



Veteran [redacted] @ [redacted] · 11h

@ [redacted] @realDonaldTrump u know what's absolutely hilarious, is I've patrolled Mosul and dealt w al queda snipers. I don't fear shit



37



Cross-Cutting Social Identities



1st Sgt. [redacted] · 11h

@ [redacted] @realDonaldTrump I spilled blood in the desert for two tours one in storm and one in shield, ure point is.



1st Sgt. [redacted] · 11h

@ [redacted] @realDonaldTrump its ure.vote. ure right. I just don't see the logic. Good luck



Veteran [redacted] · 11h

@ [redacted] @realDonaldTrump I'm happy to argue, but not to be called a disgrace. I'm not on day zero of basic



1st Sgt. [redacted] · 11h

@ [redacted] @realDonaldTrump ure opinion is ures and mine is mine. Your right tho I should of kept the fact of ure disgrace to myself



Veteran [redacted] · 11h

@ [redacted] @realDonaldTrump i don't really take it personally if it's another vet. I'm sorry i was upset



Cross-Cutting Social Identities

**1st Sgt. [redacted]** 11h
@ [redacted] @realDonaldTrump sorry for the comment
 2 ...

**Veteran [redacted]** 11h
@ [redacted] @realDonaldTrump me too. I have mad respect for you, one tour was enough
 4 ...

**1st Sgt. [redacted]** 11h
@ [redacted] @realDonaldTrump My grandfather , same division. He had it tougher though. Omaha Beach ww2
 1 ...

**Veteran [redacted]** 11h
@ [redacted] @realDonaldTrump jesus, yes.
 ...

**1st Sgt. [redacted]** 11h
@ [redacted] @realDonaldTrump These wars are gonna be the longest and most difficult to win. Its like a widesread Vietnam
 2 ...

**Veteran [redacted]** 11h
@ [redacted] @realDonaldTrump absolutely. 15 years and counting, way too much. At least there is some hope in Iraq
 2 ...



1st Sgt. [redacted]



 Follow

@ [redacted] @realDonaldTrump With the right leader and Generals in place

LIKE
1



Project Outline

- Find partisans engaged in incivil arguments on Twitter

Project Outline

- Find partisans engaged in incivil arguments on Twitter
- Use bots to send messages to ge them to stop

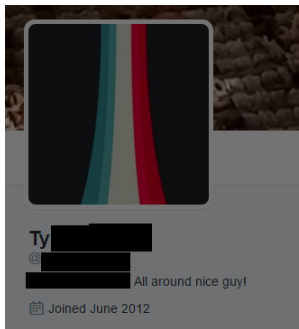
Project Outline

- Find partisans engaged in incivil arguments on Twitter
- Use bots to send messages to ge them to stop
 - ▶ Feelings tweetment
 - ▶ Rules tweetment
 - ▶ Public (placebo) tweetment

Project Outline

- Find partisans engaged in incivil arguments on Twitter
- Use bots to send messages to ge them to stop
 - ▶ Feelings tweetment
 - ▶ Rules tweetment
 - ▶ Public (placebo) tweetment
- Measure changes in directed incivility over time

Political incivility



Who Cares?

- Twitter...is important for US politics

TRUMP: Well, let me tell you about Twitter. I think that maybe I wouldn't be here if it wasn't for Twitter, because I get such a fake press, such a dishonest press.

TRUMP: Well, let me tell you about Twitter. I think that maybe I wouldn't be here if it wasn't for Twitter, because I get such a fake press, such a dishonest press.



Who Cares?

- Twitter...is important for US politics
- High levels of “affect polarization” in the offline electorate (Iyengar & Westwood, 2015)

Who Cares?

- Twitter...is important for US politics
- High levels of “affect polarization” in the offline electorate (Iyengar & Westwood, 2015)
- Exposure to incivility leads to greater affect polarization and less information acquisition (Mutz 2015)

Who Cares?

- Twitter...is important for US politics
- High levels of “affect polarization” in the offline electorate (Iyengar & Westwood, 2015)
- Exposure to incivility leads to greater affect polarization and less information acquisition (Mutz 2015)
- Mutz (2015): “uncivil discourse is communication that violates the norms of politeness of a given culture...Following the rules of civility/politeness is...a means of demonstrating mutual respect.”

Who Cares?

- Twitter...is important for US politics
- High levels of “affect polarization” in the offline electorate (Iyengar & Westwood, 2015)
- Exposure to incivility leads to greater affect polarization and less information acquisition (Mutz 2015)
- Mutz (2015): “uncivil discourse is communication that violates the norms of politeness of a given culture...Following the rules of civility/politeness is...a means of demonstrating mutual respect.”
- Incivil discourse may make deliberative democracy *impossible*

Who Cares?

- Twitter...is important for US politics
- High levels of “affect polarization” in the offline electorate (Iyengar & Westwood, 2015)
- Exposure to incivility leads to greater affect polarization and less information acquisition (Mutz 2015)
- Mutz (2015): “uncivil discourse is communication that violates the norms of politeness of a given culture...Following the rules of civility/politeness is...a means of demonstrating mutual respect.”
- Incivil discourse may make deliberative democracy *impossible*
 - ▶ Participants sincerely weigh the merits of arguments, regardless of who makes them (Fishkin, 2009)

Political incivility online

- Computer-mediated communication lacks biological feedback

Political incivility online

- Computer-mediated communication lacks biological feedback
- Competition for attention online: “It may be easy to speak in cyberspace, but it remains difficult to be heard” (Hindman, 2008)

Political incivility online

- Computer-mediated communication lacks biological feedback
- Competition for attention online: “It may be easy to speak in cyberspace, but it remains difficult to be heard” (Hindman, 2008)
- Small number of committed bad actors (“trolls”), on eg 4chan (Phillips, 2015)

Political incivility online

- Computer-mediated communication lacks biological feedback
- Competition for attention online: “It may be easy to speak in cyberspace, but it remains difficult to be heard” (Hindman, 2008)
- Small number of committed bad actors (“trolls”), on eg 4chan (Phillips, 2015)
- Compositional effects: incivility can cause people to disengage from politics on social media—even politicians (Theocharis et al, 2016)

Political incivility online

- Computer-mediated communication lacks biological feedback
- Competition for attention online: “It may be easy to speak in cyberspace, but it remains difficult to be heard” (Hindman, 2008)
- Small number of committed bad actors (“trolls”), on eg 4chan (Phillips, 2015)
- Compositional effects: incivility can cause people to disengage from politics on social media—even politicians (Theocharis et al, 2016)
- Direct effects: seeing uncivil comments can cause a wider group of people to act uncivilly (Cheng et al, 2017)

Manipulating political discourse

- Experiments in the lab
 - ▶ Convenience samples
 - ▶ Short time frame
 - ▶ In the lab

~~Manipulating political discourse~~ My Approach

- ~~Experiments in the lab~~ Experiment in the “field”
 - ▶ ~~Convenience samples~~ Sample of real, consistently uncivil users
 - ▶ ~~Short time frame~~ Continuous and unbounded time frame
 - ▶ ~~In the lab~~ In the same context as the uncivil political discussion

Finding political incivility

- Needs to be fast, and accurate

Finding political incivility

- Needs to be fast, and accurate
- Don't care about recall—but implications for external validity

Finding political incivility

- Needs to be fast, and accurate
- Don't care about recall—but implications for external validity
- Only interested in:
 - ▶ real (non-elite) users

Finding political incivility

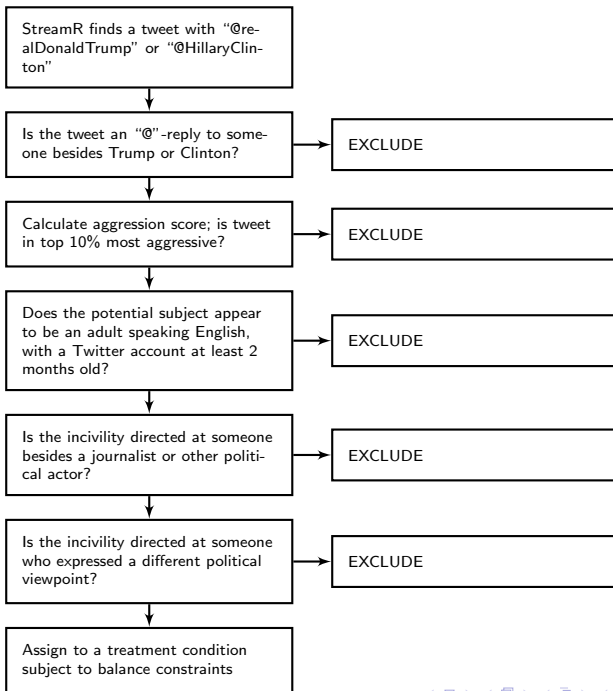
- Needs to be fast, and accurate
- Don't care about recall—but implications for external validity
- Only interested in:
 - ▶ real (non-elite) users
 - ▶ arguing with out-partisans

Finding political incivility

- Needs to be fast, and accurate
- Don't care about recall—but implications for external validity
- Only interested in:
 - ▶ real (non-elite) users
 - ▶ arguing with out-partisans
 - ▶ about politics

Finding political incivility





A Visual Overview

**Donald J. Trump** @realDonaldTrump · Oct 11
In Texas now, leaving soon for BIG rally in Florida!

 8.4K  30K ...

**NSA agent jim** @ [REDACTED] · Oct 11
@realDonaldTrump HOPE U DIE

 6  92 ...



**Sean** [REDACTED]
@ [REDACTED]







@ [REDACTED] @realDonaldTrump typical
libtard go fuck yourself

11:25 AM - 12 Oct 2016

 ...

 @realDonaldTrump

**Matthew** [REDACTED] @Matthew [REDACTED] Oct 12 **FEELINGS**
@ [REDACTED] You shouldn't use language like that. Republicans need to remember that our opponents are real people, with real feelings.

 ...

Measuring incivility

- Use a machine learning model to evaluate “aggressiveness” in Wikipedia comments (Wulczyn, Thain & Dixon, 2017)

Measuring incivility

- Use a machine learning model to evaluate “aggressiveness” in Wikipedia comments (Wulczyn, Thain & Dixon, 2017)
- Trained on over 100,000 human-coded comments

Measuring incivility

- Use a machine learning model to evaluate “aggressiveness” in Wikipedia comments (Wulczyn, Thain & Dixon, 2017)
- Trained on over 100,000 human-coded comments
- Model is a multi-layer perceptron using character-level n-grams; extremely black box, but more accurate than a single coder

Measuring incivility

- Use a machine learning model to evaluate “aggressiveness” in Wikipedia comments (Wulczyn, Thain & Dixon, 2017)
- Trained on over 100,000 human-coded comments
- Model is a multi-layer perceptron using character-level n-grams; extremely black box, but more accurate than a single coder
- Pre-registered, not related to the existing data

Measuring incivility

- Use a machine learning model to evaluate “aggressiveness” in Wikipedia comments (Wulczyn, Thain & Dixon, 2017)
- Trained on over 100,000 human-coded comments
- Model is a multi-layer perceptron using character-level n-grams; extremely black box, but more accurate than a single coder
- Pre-registered, not related to the existing data
- Restricted subject tweets (367k) to those directed “@” another user

Measuring incivility

- Use a machine learning model to evaluate “aggressiveness” in Wikipedia comments (Wulczyn, Thain & Dixon, 2017)
- Trained on over 100,000 human-coded comments
- Model is a multi-layer perceptron using character-level n-grams; extremely black box, but more accurate than a single coder
- Pre-registered, not related to the existing data
- Restricted subject tweets (367k) to those directed “@” another user
- Classified a tweet as a uncivil if score $>$ 75th percentile (robust to 70th or 80th)

Tweetment Variations

- Moral intuitionist model (Haidt, 2001): moral emotion is antecedent to moral reasoning

Tweetment Variations

- Moral intuitionist model (Haidt, 2001): moral emotion is antecedent to moral reasoning
- Moral appeals should target intuitions rather than (epiphenominal) logic

Tweetment Variations

- Moral intuitionist model (Haidt, 2001): moral emotion is antecedent to moral reasoning
- Moral appeals should target intuitions rather than (epiphenominal) logic
- Different rhetoric to appeal to different moral frameworks

Tweetment Variations

- Moral intuitionist model (Haidt, 2001): moral emotion is antecedent to moral reasoning
- Moral appeals should target intuitions rather than (epiphenominal) logic
- Different rhetoric to appeal to different moral frameworks
 - ▶ Authority moral foundation: “You shouldn’t use language like that. [Democrats/Republicans] need to behave according to the proper rules of political civility.”

Tweetment Variations

- Moral intuitionist model (Haidt, 2001): moral emotion is antecedent to moral reasoning
- Moral appeals should target intuitions rather than (epiphenominal) logic
- Different rhetoric to appeal to different moral frameworks
 - ▶ Authority moral foundation: “You shouldn’t use language like that. [Democrats/Republicans] need to behave according to the proper rules of political civility.”
 - ▶ Care moral foundation: “You shouldn’t use language like that. [Democrats/Republicans] need to remember that our opponents are real people, with real feelings.”

Tweetment Variations

- Moral intuitionist model (Haidt, 2001): moral emotion is antecedent to moral reasoning
- Moral appeals should target intuitions rather than (epiphenominal) logic
- Different rhetoric to appeal to different moral frameworks
 - ▶ Authority moral foundation: “You shouldn’t use language like that. [Democrats/Republicans] need to behave according to the proper rules of political civility.”
 - ▶ Care moral foundation: “You shouldn’t use language like that. [Democrats/Republicans] need to remember that our opponents are real people, with real feelings.”
 - ▶ Placebo message: “Remember that everything you post here is public. Everyone can see that you tweeted this.”

Tweetment uptake

The image shows a Twitter profile for 'Hillary for President'. The banner at the top features the word 'Hillary' in large white letters on a blue background, with 'for President' in a smaller script font. Below the banner is the user's profile picture, a cartoon avatar of a man. To the right of the profile picture are statistics: TWEETS 183, FOLLOWING 40, FOLLOWERS 922, and MOMENTS 0. Below these statistics is an 'Edit profile' button. The profile name is 'Hillary 2016!' and the bio is 'New York, NY'. The user joined in January 2015. The 'Tweets' tab is selected, showing a tweet from 'NYCT Subway' about a service alert. The 'Who to follow' section lists 'Jamie Stuber' and 'Washington Post'. The 'Trends' section shows '#CMYyourCrushDay'.

Hillary for President

TWEETS 183 FOLLOWING 40 FOLLOWERS 922 MOMENTS 0

Edit profile

Neil
@Neil
Hillary 2016!
New York, NY
Joined January 2015

Photos and videos

Tweets Tweets & replies Media

You Retweeted
NYCT Subway @NYCTSubway · Oct 20
#ServiceAlert: Following an earlier incident at Queens Plaza, E and M train service has resumed with delays.
2 3

You Retweeted
Goings On About Town @goingson · Oct 19
From Gujarat to Houston to New York, Pondicheri brings Nirvanic Indian food to the Flatiron district.

Who to follow · Refresh · View all

Jamie Stuber @JamieSt...
Follow

Washington Post @was...
Follow

Find friends

Trends · Change
#CMYyourCrushDay
259K Tweets

Tweetment uptake

The screenshot shows a Twitter profile for a user named Todd (@Todd). The profile picture is a cartoon illustration of a man with dark hair. The background image is a large red elephant, a symbol of the Republican Party, with the word "REPUBLICAN" written in white capital letters on its side. The profile statistics show 192 tweets, 40 following, 933 followers, and 0 moments. The bio mentions "Republicans 2016" and "New York, NY". The "Tweets" tab is selected, showing a tweet from Grub Street (@grubstreet) dated Oct 20. The tweet text is "The sad tale of a restaurant's 'pay what you want' policy failing spectacularly" and includes a photo of a red table with some cash and coins. The right sidebar shows "Who to follow" with Kris Sanchez and Rihanna, and "Trends" for #DMYourCrushDay and #CNY16.

Todd
@Todd
Republicans 2016
New York, NY
Joined January 2015

TWEETS 192 FOLLOWING 40 FOLLOWERS 933 MOMENTS 0

Edit profile

Tweets Tweets & replies Media

You Retweeted
Grub Street @grubstreet · Oct 20
The sad tale of a restaurant's "pay what you want" policy failing spectacularly

Who to follow · Refresh · View all
Kris Sanchez @KrisSan...
Follow
Rihanna @rihanna
Follow
Find trends

Trends · Change
#DMYourCrushDay
274K Tweets
#CNY16

Hypotheses

Hypothesis pre-registered through EGAP.

Hypothesis

The effect of the Care condition will be larger for liberals than for conservatives. There will be an effect of the Authority condition for conservatives, but not for liberals. There will be an effect of the Public condition, but it will be smaller than the other effects.

Hypotheses

Hypothesis pre-registered through EGAP.

Hypothesis

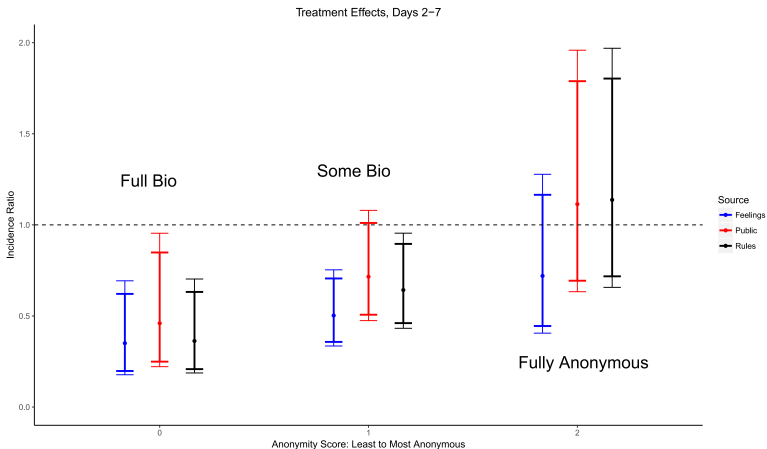
The effect of the Care condition will be larger for liberals than for conservatives. There will be an effect of the Authority condition for conservatives, but not for liberals. There will be an effect of the Public condition, but it will be smaller than the other effects.

Not pre-registered; derived from Munger (2016).

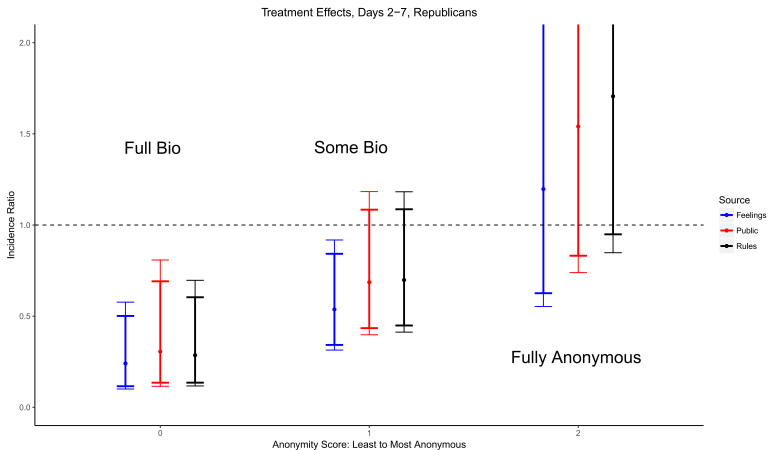
Hypothesis

Treatment effects will be larger for more anonymous subjects.

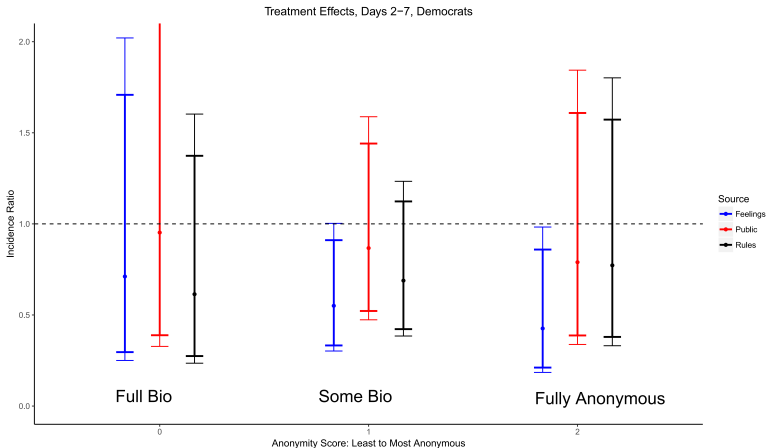
Change in Incivility: One Week Post-Treatment (N = 310)



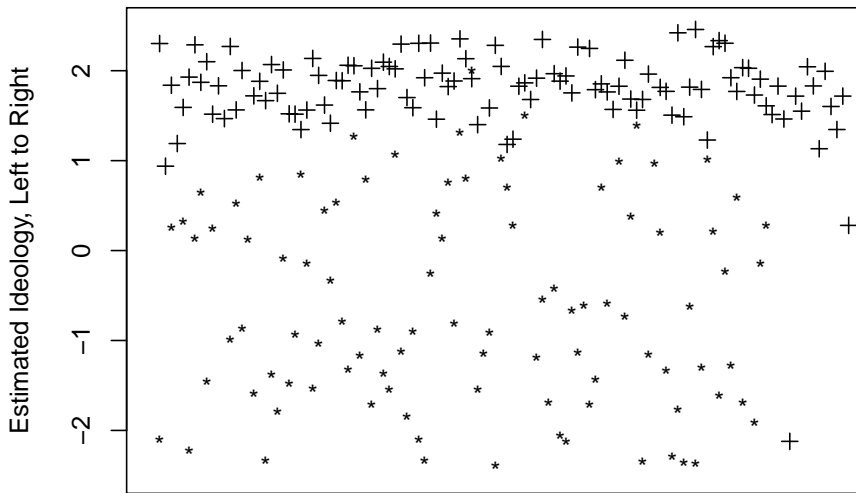
Change in Incivility: Republicans (N=163)



Change in Incivility: Democrats (N=147)



Estimated v Labeled Ideology



Restricting to “True Liberals”

Feelings Treatment Effects

	All Liberals (N=147)	True Liberals (N=86)
1 Day	-.63	-.84
2-7 Days	-.33	-.63
15-28 Days	.20	-.05

Rules Treatment Effects

	All Liberals (N=147)	True Liberals (N=86)
1 Day	-.77	-1.06
2-7 Days	-.54	-1.27
15-28 Days	-.71	-1.04

Public Treatment Effects

	All Liberals (N=147)	True Liberals (N=86)
1 Day	.52	.85
2-7 Days	-.06	-.08
15-28 Days	.25	.23

Encouraging Online Civility

- Large and persistent treatment effects (on most of the sample) from a single intervention

Encouraging Online Civility

- Large and persistent treatment effects (on most of the sample) from a single intervention
- No evidence of differential effectiveness of variations in moral suasion

Encouraging Online Civility

- Large and persistent treatment effects (on most of the sample) from a single intervention
- No evidence of differential effectiveness of variations in moral suasion
- Small, committed groups promoting incivility online

Encouraging Online Civility

- Large and persistent treatment effects (on most of the sample) from a single intervention
- No evidence of differential effectiveness of variations in moral suasion
- Small, committed groups promoting incivility online
 - ▶ Trolls
 - ▶ Ideologues

Encouraging Online Civility

- Large and persistent treatment effects (on most of the sample) from a single intervention
- No evidence of differential effectiveness of variations in moral suasion
- Small, committed groups promoting incivility online
 - ▶ Trolls
 - ▶ Ideologues
- My hope: most people would prefer civility

Thanks for your comments, and for listening!

km2713@nyu.edu
@kmmunger (please be civil)

Negative Binomial Specification

$$\begin{aligned} \ln(Agg_{post}) = & x_{int} + \beta_1 Agg_{pre} + \beta_2 T_{feel} + \beta_3 T_{rules} + \beta_4 T_{public} + \\ & \beta_5 Anon + \beta_6 (T_{feel} \times Anon) + \\ & \beta_7 (T_{rules} \times Anon) + \beta_8 (T_{public} \times Anon) \end{aligned}$$

$$IRR_{feel \times Anon_1} = e^{\hat{\beta}_2 + \hat{\beta}_6 \times 1}$$

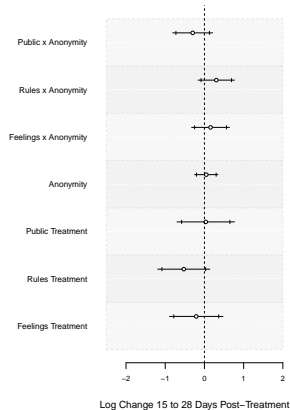
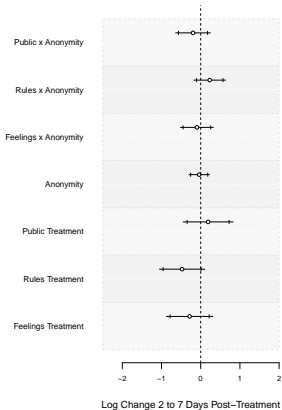
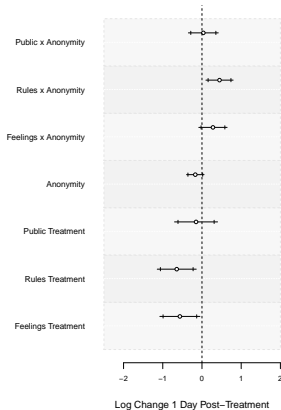
$$V_{feel \times Anon_1} = V(\hat{\beta}_2) + Anon^2 V(\hat{\beta}_6) + 2Anon \times Cov(\hat{\beta}_2, \hat{\beta}_6)$$

Attrition rates

Table: Attrition Rates and Causes

	Control	Liberals	Conservatives
Initial assignment	108	104	118
Failed treatment application	0	2	2
Tweeted too often/bots	3	1	5
Suspended	0	1	2
Weird	2	0	0
Final	102	100	108
Attrition	6%	4%	8%

Full Sample: OLS (N=310)



Excluding Mislabeled Liberals: OLS (N=249)

