Machine Learning

Justin Grimmer

University of Chicago

March 5th, 2018

Causal Inference and Text

- Text as Treatment
- Text as Outcome
- Text as Confounder

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Text as Outcome

- Text is a rich source of information about the opinions, views and responses of individuals.
- Most instances so far in political science of people collecting large text datasets have been text as outcome
- Also includes a long history of manual coding of open-ended survey responses and manual content analysis of documents.
- We will define our estimand as:

$$E[g(Y_i(1)) - g(Y_i(0))]$$

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Designing a g Function

- Before we can talk about learning the g function, need to talk about desirable properties.
- Desirable properties of g function
  1. Interpretable
     can we clearly communicate the idea to the reader
  2. Theoretical Interest
     helps us advance a relevant argument
  3. Label Fidelity
     minimal surprise when going from reading the label to reading the documents
  4. Tractable
     computationally tractable model and enough samples to estimate
- We will consider unsupervised learning of g function (works because unsupervised learning does dimensionality reduction)

Types of g Functions

The biggest modeling choice is the form of the latent representation.
There are many options:

- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Types of g Functions

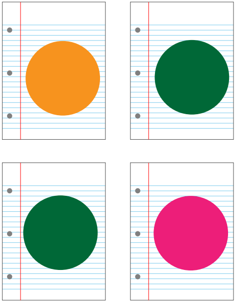
The biggest modeling choice is the form of the latent representation. There are many options:

- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Types of g Functions

The biggest modeling choice is the form of the latent representation. There are many options:



- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Types of g Functions

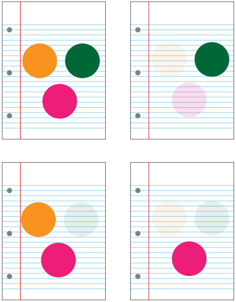
The biggest modeling choice is the form of the latent representation. There are many options:

- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Types of g Functions

The biggest modeling choice is the form of the latent representation. There are many options:

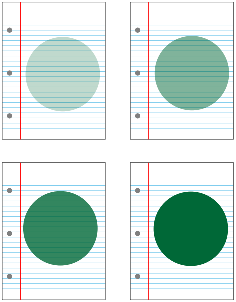


- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Types of g Functions

The biggest modeling choice is the form of the latent representation.
There are many options:



- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Types of g Functions

The biggest modeling choice is the form of the latent representation. There are many options:

- Categorical: one of K mutually exclusive and exhaustive categories
- Mixed Membership: proportional member of K topics
- Binary Features: K binary latent variables, each of which could be one or off
- Scales: K continuous scales or positions

We are going to use the Structural Topic Model (Roberts, Stewart and Tingley) which provides a mixed membership representation, but the exact method isn't important.

Threats to Inference

- If we have a g function before seeing any documents we have no problems.
- If not, how we discover the g function is important.
- When we use the same documents to discover g (by manual or automated means) and estimate the treatment effect, we induce a dependence across all observations.
- Now $g(Y_i(\mathbf{T}))$ depends on all elements of $\mathbf{T}$ because treatment assignment of all documents affected our development of g.
- We call this problem an Analyst Induced SUTVA Violation.

Threats to Inference

- If we have a g function before seeing any documents we have no problems.
- If not, how we discover the g function is important.
- When we use the same documents to discover g (by manual or automated means) and estimate the treatment effect, we induce a dependence across all observations.
- Now $g(Y_i(\mathbf{T}))$ depends on all elements of $\mathbf{T}$ because treatment assignment of all documents affected our development of g.
- We call this problem an Analyst Induced SUTVA Violation.

Threats to Inference

- If we have a g function before seeing any documents we have no problems.
- If not, how we discover the g function is important.
- When we use the same documents to discover g (by manual or automated means) and estimate the treatment effect, we induce a dependence across all observations.
- Now $g(Y_i(\mathbf{T}))$ depends on all elements of $\mathbf{T}$ because treatment assignment of all documents affected our development of g.
- We call this problem an Analyst Induced SUTVA Violation.

Threats to Inference

- If we have a g function before seeing any documents we have no problems.
- If not, how we discover the g function is important.
- When we use the same documents to discover g (by manual or automated means) and estimate the treatment effect, we induce a dependence across all observations.
- Now $g(Y_i(\mathbf{T}))$ depends on all elements of $\mathbf{T}$ because treatment assignment of all documents affected our development of g.
- We call this problem an Analyst Induced SUTVA Violation.

Threats to Inference

- If we have a g function before seeing any documents we have no problems.
- If not, how we discover the g function is important.
- When we use the same documents to discover g (by manual or automated means) and estimate the treatment effect, we induce a dependence across all observations.
- Now $g(Y_i(\mathbf{T}))$ depends on all elements of $\mathbf{T}$ because treatment assignment of all documents affected our development of g.
- We call this problem an Analyst Induced SUTVA Violation.

Threats to Inference

- If we have a g function before seeing any documents we have no problems.
- If not, how we discover the g function is important.
- When we use the same documents to discover g (by manual or automated means) and estimate the treatment effect, we induce a dependence across all observations.
- Now $g(Y_i(\mathbf{T}))$ depends on all elements of $\mathbf{T}$ because treatment assignment of all documents affected our development of g.
- We call this problem an Analyst Induced SUTVA Violation.

Avoiding Analyst Induced SUTVA Violations

AISVs are pernicious because they are exacerbated by what would otherwise be best practice.

Consider hand-coding: when we iterate between writing the codebook, classifying statements and analyzing intercoder agreement, we induce dependence.

We can avoid the AISV with a heroic assumption that the codebook (g function) doesn't depend on the specific randomization, i.e. that g is stable across all randomizations.

While this avoids the AISV, it doesn't remove concerns about fishing.

Avoiding Analyst Induced SUTVA Violations

AISVs are pernicious because they are exacerbated by what would otherwise be best practice.

Consider hand-coding: when we iterate between writing the codebook, classifying statements and analyzing intercoder agreement, we induce dependence.

We can avoid the AISV with a heroic assumption that the codebook (g function) doesn't depend on the specific randomization, i.e. that g is stable across all randomizations.

While this avoids the AISV, it doesn't remove concerns about fishing.

Avoiding Analyst Induced SUTVA Violations

AISVs are pernicious because they are exacerbated by what would otherwise be best practice.

Consider hand-coding: when we iterate between writing the codebook, classifying statements and analyzing intercoder agreement, we induce dependence.

We can avoid the AISV with a heroic assumption that the codebook (g function) doesn't depend on the specific randomization, i.e. that g is stable across all randomizations.

While this avoids the AISV, it doesn't remove concerns about fishing.

Avoiding Analyst Induced SUTVA Violations

AISVs are pernicious because they are exacerbated by what would otherwise be best practice.

Consider hand-coding: when we iterate between writing the codebook, classifying statements and analyzing intercoder agreement, we induce dependence.

We can avoid the AISV with a heroic assumption that the codebook (g function) doesn't depend on the specific randomization, i.e. that g is stable across all randomizations.

While this avoids the AISV, it doesn't remove concerns about fishing.

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

A General Solution: Train-Test Splits

- We can address this problem by explicitly allowing for discovery in the research process.
- Randomly partition sample into two sets: training and test
- Training Set: do whatever we want to find the best g function (useful and provides peace of mind).
- Test Set: Estimate causal effects using the learned g.
- This addresses:
  - AISV: g does not depend on randomization in test set.
  - Overfitting: any fishing in training set will not produce result in test set.

Not coincidentally there is new functionality in the `stm` package that allows you to apply the g function to a test set

Tradeoffs With Train-Test Splits

- Train-test split is not costless: biggest concern is power.
- It can be challenging to set the train-test split ratio because we don't know the power we need for discovery
- We might also want to know that our discovery is invariant to the train-test split- although we note that it isn't strictly necessary.
- Also relies on the premise of well-intentioned actors who aren't 'peeking' at the test set (although with outsourced data collection, some ways around this)

We lose some power, but gain a process that preserves identification of the causal effect.

Tradeoffs With Train-Test Splits

- Train-test split is not costless: biggest concern is power.
- It can be challenging to set the train-test split ratio because we don't know the power we need for discovery
- We might also want to know that our discovery is invariant to the train-test split- although we note that it isn't strictly necessary.
- Also relies on the premise of well-intentioned actors who aren't 'peeking' at the test set (although with outsourced data collection, some ways around this)

We lose some power, but gain a process that preserves identification of the causal effect.

Tradeoffs With Train-Test Splits

- Train-test split is not costless: biggest concern is power.
- It can be challenging to set the train-test split ratio because we don't know the power we need for discovery
- We might also want to know that our discovery is invariant to the train-test split- although we note that it isn't strictly necessary.
- Also relies on the premise of well-intentioned actors who aren't 'peeking' at the test set (although with outsourced data collection, some ways around this)

We lose some power, but gain a process that preserves identification of the causal effect.

Tradeoffs With Train-Test Splits

- Train-test split is not costless: biggest concern is power.
- It can be challenging to set the train-test split ratio because we don't know the power we need for discovery
- We might also want to know that our discovery is invariant to the train-test split- although we note that it isn't strictly necessary.
- Also relies on the premise of well-intentioned actors who aren't 'peeking' at the test set (although with outsourced data collection, some ways around this)

We lose some power, but gain a process that preserves identification of the causal effect.

Tradeoffs With Train-Test Splits

- Train-test split is not costless: biggest concern is power.
- It can be challenging to set the train-test split ratio because we don't know the power we need for discovery
- We might also want to know that our discovery is invariant to the train-test split- although we note that it isn't strictly necessary.
- Also relies on the premise of well-intentioned actors who aren't 'peeking' at the test set (although with outsourced data collection, some ways around this)

We lose some power, but gain a process that preserves identification of the causal effect.

Tradeoffs With Train-Test Splits

- Train-test split is not costless: biggest concern is power.
- It can be challenging to set the train-test split ratio because we don't know the power we need for discovery
- We might also want to know that our discovery is invariant to the train-test split- although we note that it isn't strictly necessary.
- Also relies on the premise of well-intentioned actors who aren't 'peeking' at the test set (although with outsourced data collection, some ways around this)

We lose some power, but gain a process that preserves identification of the causal effect.

Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

Immigration Application: Experiment 1

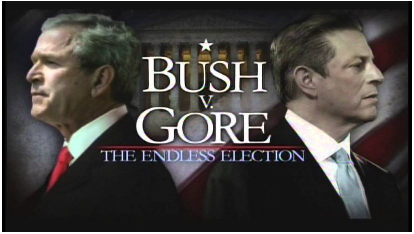
- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.



Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

> "A 28-year-old single man, a citizen of another country, was convicted of illegally entering the United States. Prior to this offense, he had served two previous prison sentences each more than a year. One of these previous sentences was for a violent crime and he had been deported back to his home country."

Immigration Application: Experiment 1

- Example application on a survey experiment about attitudes toward immigration.
- Uses data from a study by Cohen, Rust and Steen (2004), telephone random-digit dial of 1300 respondents (conducted in 2000). Train: 50%, Test 50%.
- Respondents given a prompt about an immigrant, asked if she should be imprisoned and *if they say no* why.

> "A 28-year-old single man, a citizen of another country, was convicted of illegally entering the United States. Prior to this offense, he had never been imprisoned before."

Immigration Application: Experiment 2

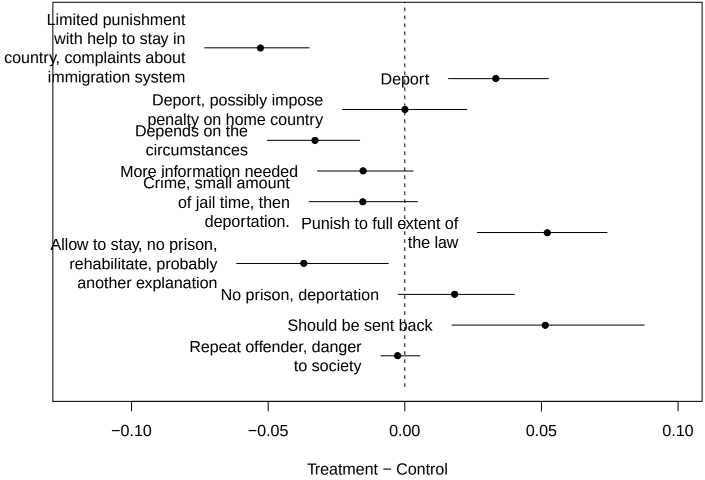
- Update sample using contemporaneous participants
- Alter the prompt: "Should this offender be sent to prison?" (responses: yes, no, don't know) $\rightsquigarrow$ "Why or Why not? Please describe in at least two sentences what actions, if any, the US government should take with respect to this person and why"

Immigration Application: Experiment 3

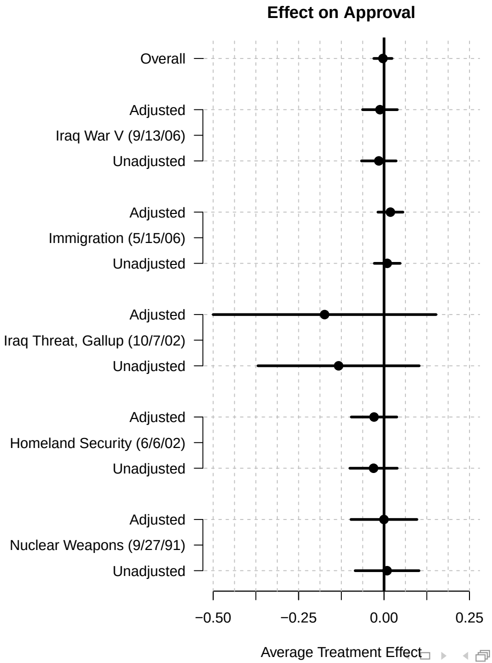
- Examining Experiment 2: we noticed our labels were poorly constructed
- Cannot revise labels!
- Rerun experiment, team label the output

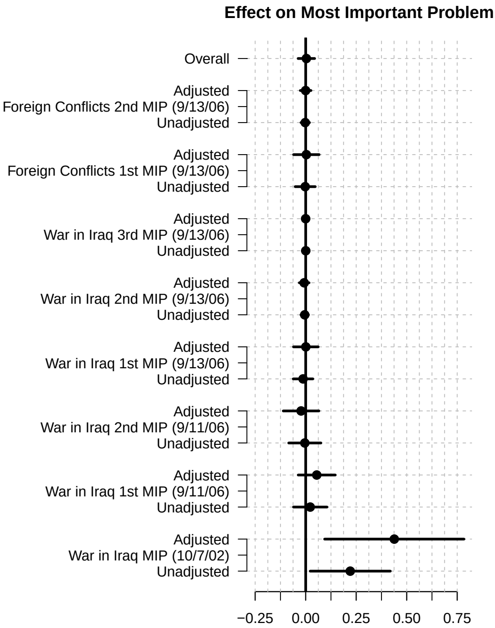
	Label	Highest Probability Words
Topic 1	Limited punishment with help to stay in country, complaints about immigration system	legal, way, immigr, danger, peopl, allow, come, countri, can, enter
Topic 2	Deport	deport, think, prison, crime, alreadi, imprison, illeg, sinc, serv, time
Topic 3	Deport because of money	just, send, back, countri, jail, come, prison, let, harm, money
Topic 4	Depends on the circumstances	first, countri, time, came, jail, man, think, reason, govern, put
Topic 5	More information needed	state, unit, prison, crime, immigr, illeg, take, crimin, simpli, put
Topic 6	Crime, small amount of jail time, then deportation	enter, countri, illeg, person, jail, deport, time, proper, imprison, determin
Topic 7	Punish to full extent of the law	crime, violent, person, law, convict, commit, deport, illeg, punish, offend
Topic 8	Allow to stay, no prison, rehabilitate, probably another explanation	dont, crimin, think, tri, hes, offens, better, case, know, make
Topic 9	No prison, deportation	deport, prison, will, person, countri, man, illeg, serv, time, sentenc
Topic 10	Should be sent back	sent, back, countri, prison, home, think, pay, origin, illeg, time
Topic 11	Repeat offender, danger to society	believ, countri, violat, offend, person, law, deport, prison, citizen, individu

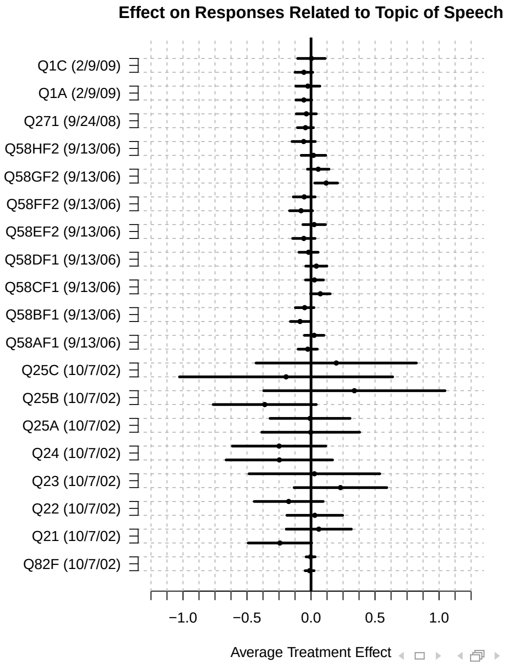
$$\widehat{ATE} = \sum_{i \in \boldsymbol{I}} \frac{I(T_i = 1) g_{\boldsymbol{K}}(\boldsymbol{Y}_i(1))}{\sum_{i \in \boldsymbol{I}} I(T_i = 1)} - \sum_{i \in \boldsymbol{I}} \frac{I(T_i = 0) g_{\boldsymbol{K}}(\boldsymbol{Y}_i(0))}{\sum_{i \in \boldsymbol{I}} I(T_i = 0)}$$


Limited punishment with help to stay in country, complaints about immigration system
Deport
Deport, possibly impose penalty on home country
Depends on the circumstances
More information needed
Crime, small amount of jail time, then deportation.
Punish to full extent of the law
Allow to stay, no prison, rehabilitate, probably another explanation
No prison, deportation
Should be sent back
Repeat offender, danger to society
−0.10
−0.05
0.00
0.05
0.10
Treatment − Control

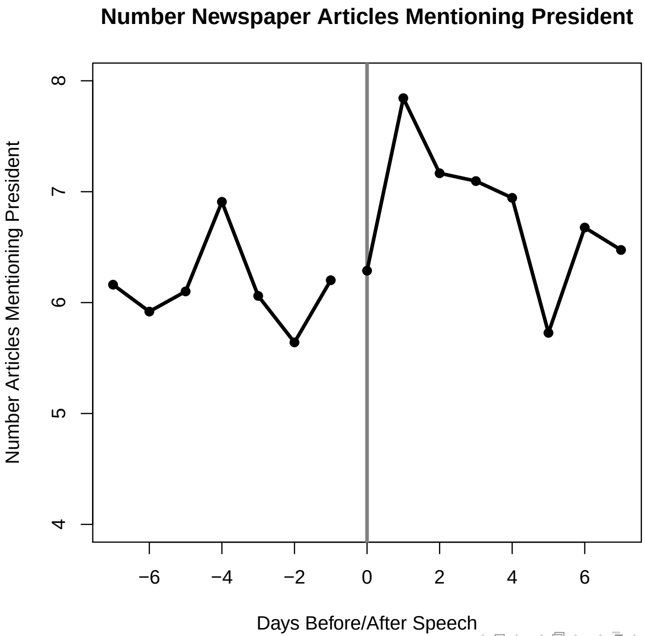
How do presidents “going public" affect public opinion?


Effect on Approval
Overall
Adjusted
Iraq War V (9/13/06)
Unadjusted
Adjusted
Immigration (5/15/06)
Unadjusted
Adjusted
Iraq Threat, Gallup (10/7/02)
Unadjusted
Adjusted
Homeland Security (6/6/02)
Unadjusted
Adjusted
Nuclear Weapons (9/27/91)
Unadjusted
−0.50
−0.25
0.00
0.25
Average Treatment Effect


Effect on Most Important Problem
Overall
Adjusted
Foreign Conflicts 2nd MIP (9/13/06)
Unadjusted
Adjusted
Foreign Conflicts 1st MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 3rd MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 2nd MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 1st MIP (9/13/06)
Unadjusted
Adjusted
War in Iraq 2nd MIP (9/11/06)
Unadjusted
Adjusted
War in Iraq 1st MIP (9/11/06)
Unadjusted
Adjusted
War in Iraq MIP (10/7/02)
Unadjusted
−0.25
0.00
0.25
0.50
0.75


Effect on Responses Related to Topic of Speech
Q1C (2/9/09)
Q1A (2/9/09)
Q271 (9/24/08)
Q58HF2 (9/13/06)
Q58GF2 (9/13/06)
Q58FF2 (9/13/06)
Q58EF2 (9/13/06)
Q58DF1 (9/13/06)
Q58CF1 (9/13/06)
Q58BF1 (9/13/06)
Q58AF1 (9/13/06)
Q25C (10/7/02)
Q25B (10/7/02)
Q25A (10/7/02)
Q24 (10/7/02)
Q23 (10/7/02)
Q22 (10/7/02)
Q21 (10/7/02)
Q82F (10/7/02)
−1.0
−0.5
0.0
0.5
1.0
Average Treatment Effect

How do presidents “going public" affect ~~public opinion~~ the media agenda?


Number Newspaper Articles Mentioning President
Number Articles Mentioning President
4
5
6
7
8
−6
−4
−2
0
2
4
6
Days Before/After Speech

1) (Assume) random assignment of treatments (use an interrupted time series design)

1) (Assume) random assignment of treatments (use an interrupted time series design)
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

1) (Assume) random assignment of treatments (use an interrupted time series design)
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

Function $\boldsymbol{g}$ now uncovers latent features of response: map from text to small number of categories

1) (Assume) random assignment of treatments (use an interrupted time series design)
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

Function g now uncovers latent features of response: map from text to small number of categories

$$\text{ATE}_k \ = \ \text{E}[g(\boldsymbol{Y}(1))_k - g(\boldsymbol{Y}(0))_k]$$

Discovering (Estimating) Dependent Variable

- Numerous options to discover: (hand coding, supervised models, unsupervised models, mixture)

Discovering (Estimating) Dependent Variable

- Numerous options to discover: (hand coding, supervised models, unsupervised models, mixture)
- All have same worries: (1) Analyst Induced SUTVA violation (2) Overfitting (potentially via Fishing)

Discovering (Estimating) Dependent Variable

- Numerous options to discover: (hand coding, supervised models, unsupervised models, mixture)
- All have same worries: (1) Analyst Induced SUTVA violation (2) Overfitting (potentially via Fishing)

Train/Test Split

1) (Assume) random assignment of treatments

2) Obtain text based response $\boldsymbol{Y}_i(T_i)$

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration
5) In test set:

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration
5) In test set:
    a) Infer dependent variables (using newly available updates to `STM` software (Roberts, Stewart, and Tingley 2017))

1) (Assume) random assignment of treatments
2) Obtain text based response $\boldsymbol{Y}_i(T_i)$
3) Randomly split response and text into train/test split
4) In training set: discover latent dependent variables
    a) Apply Structural Topic Model (Roberts, Stewart, and Airoldi 2017)
    b) Make final model pick based on quantitative model fit and exploration
5) In test set:
    a) Infer dependent variables (using newly available updates to `STM` software (Roberts, Stewart, and Tingley 2017))
    b) Estimate effect of treatments on topic prevalence across categories

A President's effect on newspaper agenda

A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech

A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published

A President's effect on newspaper agenda

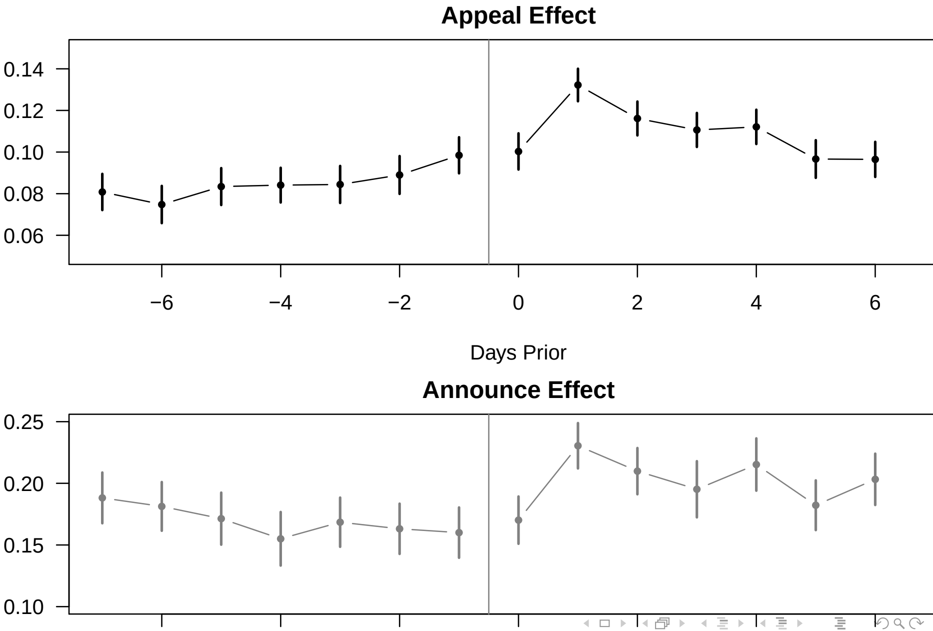
- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published
- 159,217 articles

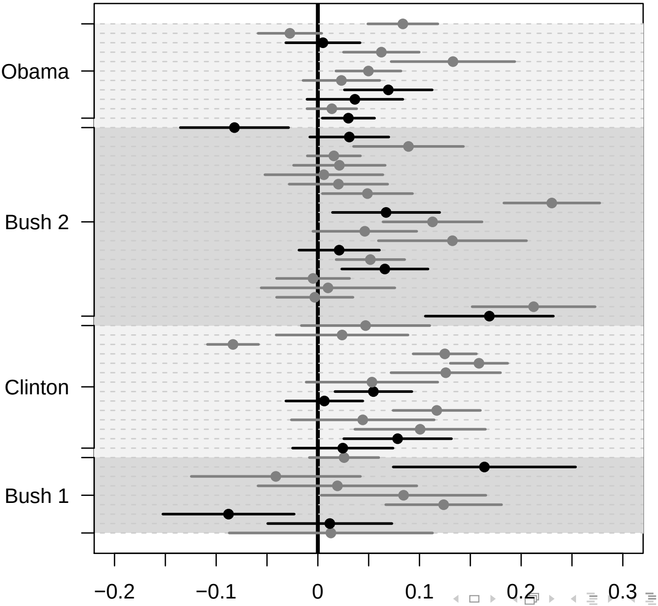
A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published
- 159,217 articles
- Train: 10%, Test 90%

A President's effect on newspaper agenda

- Response: newspaper articles mentioning `president` in 10 highest circulation papers, two-week window around speech
- Treatment: Number of days before/after speech article was published
- 159,217 articles
- Train: 10%, Test 90%
- Effect estimate: interrupted time series design on topic prevalence (compare share immediately before to share day after)


Appeal Effect
0.14
0.12
0.10
0.08
0.06
−6
−4
−2
0
2
4
6
Days Prior
Announce Effect
0.25
0.20
0.15
0.10


Obama
Bush 2
Clinton
Bush 1
−0.2
−0.1
0
0.1
0.2
0.3

Text as Confounder

Selection on Observables

Assumption:

Random Assignment: $T_i \perp\!\!\!\perp Y_i(0), Y_i(1)$

Selection on Observables: $T_i \perp\!\!\!\perp Y_i(0), Y_i(1) | \boldsymbol{X}$

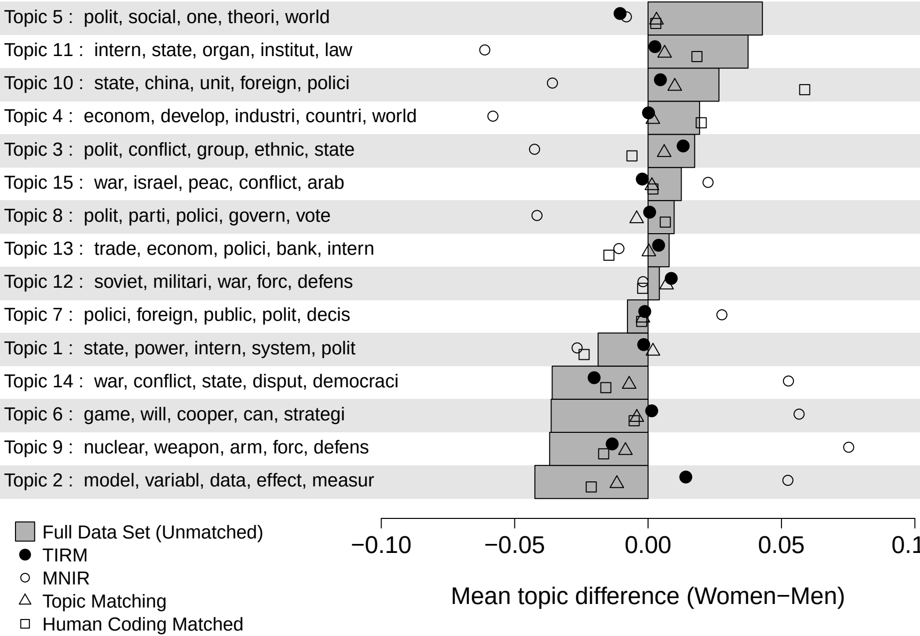
Text may be a confounder:

- Women are cited less in IR $\rightsquigarrow$ write about different subjects?
- Chinese censorship increases blogging rates $\rightsquigarrow$ systematic differences in what is censored?
- Radical cleric dying increases popularity of writing $\rightsquigarrow$ clerics targeted based on what they write?

Selection on Observables: $T_i \perp\!\!\!\perp Y_i(0), Y_i(1) | g(\boldsymbol{X})$

Nielsen, Roberts, and Stewart(2018)

- Maliniak, Powers, and Walter (2013)$\rightsquigarrow$ gender citation bias in IR
- Causal question: *same* article with man's name, different citation patterns?
- NRS: use DFR from JSTOR $\rightsquigarrow$ 3,201 IR articles, 333 by women solo(!!!!)
- Match using STM: estimate topics, coarsen exact matching, and then trimmed sample. (Use other matching procedures as well)


Topic 5 : polit, social, one, theori, world
Topic 11 : intern, state, organ, institut, law
Topic 10 : state, china, unit, foreign, polici
Topic 4 : econom, develop, industri, countri, world
Topic 3 : polit, conflict, group, ethnic, state
Topic 15 : war, israel, peac, conflict, arab
Topic 8 : polit, parti, polici, govern, vote
Topic 13 : trade, econom, polici, bank, intern
Topic 12 : soviet, militari, war, forc, defens
Topic 7 : polici, foreign, public, polit, decis
Topic 1 : state, power, intern, system, polit
Topic 14 : war, conflict, state, disput, democraci
Topic 6 : game, will, cooper, can, strategi
Topic 9 : nuclear, weapon, arm, forc, defens
Topic 2 : model, variabl, data, effect, measur
Full Data Set (Unmatched)
TIRM
MNIR
Topic Matching
Human Coding Matched
−0.10
−0.05
0.00
0.05
0.1
Mean topic difference (Women−Men)

Bigger effect: 16 fewer citations for female articles. (Naïve difference is 7)

Constituent Badges

Constituent Badges

- Visually indicates that commenter is a constituent

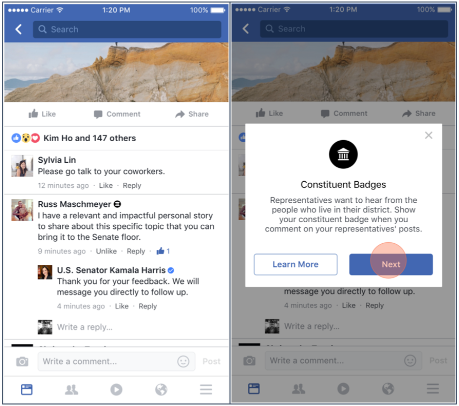
Constituent Badges

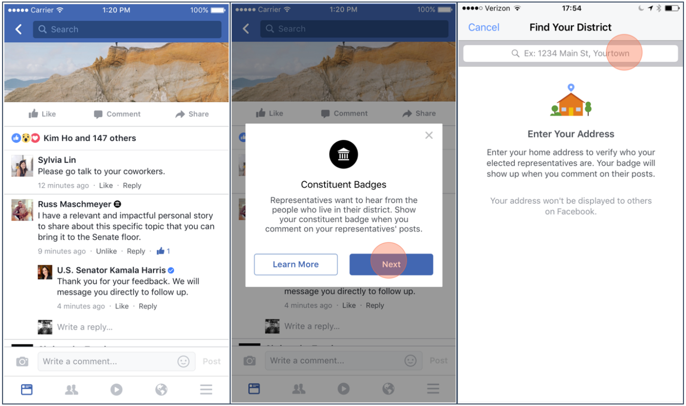
- Visually indicates that commenter is a constituent
- Inspired by Capitol Hill focus groups

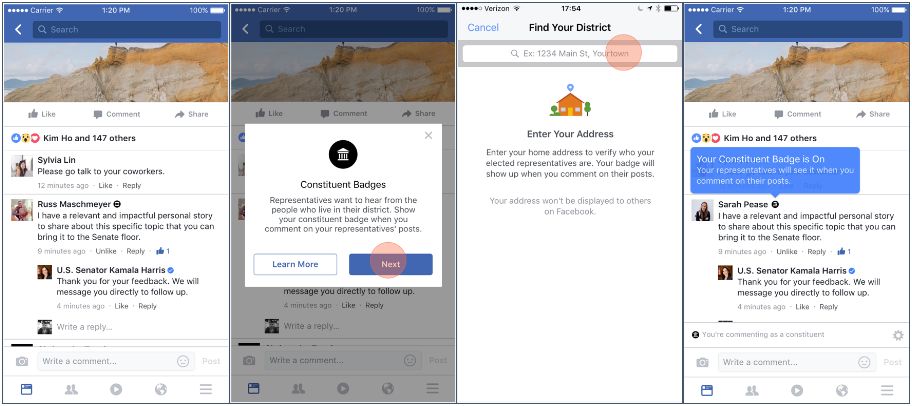
Constituent Badges

- Visually indicates that commenter is a constituent
- Inspired by Capitol Hill focus groups
- Survey responses from staff: identify constituents, we'll be responsive


●●●●● Carrier
1:20 PM
100%
Search
Like
Comment
Share
Kim Ho and 147 others
Sylvia Lin
Please go talk to your coworkers.
12 minutes ago · Like · Reply
Russ Maschmeyer
I have a relevant and impactful personal story to share about this specific topic that you can bring it to the Senate floor.
9 minutes ago · Unlike · Reply · 1
U.S. Senator Kamala Harris
Thank you for your feedback. We will message you directly to follow up.
4 minutes ago · Like · Reply
Write a reply...
Write a comment...
Post


●●●●● Carrier
1:20 PM
100%
Search
Like
Comment
Share
Kim Ho and 147 others
Sylvia Lin
Please go talk to your coworkers.
12 minutes ago · Like · Reply
Russ Maschmeyer
I have a relevant and impactful personal story to share about this specific topic that you can bring it to the Senate floor.
9 minutes ago · Unlike · Reply · 1
U.S. Senator Kamala Harris
Thank you for your feedback. We will message you directly to follow up.
4 minutes ago · Like · Reply
Write a reply...
Write a comment...
Post
Constituent Badges
Representatives want to hear from the people who live in their district. Show your constituent badge when you comment on your representatives' posts.
Learn More
Next
message you directly to follow up.
4 minutes ago · Like · Reply
Write a reply...
Write a comment...
Post


Carrier 1:20 PM 100%
Search
Like Comment Share
Kim Ho and 147 others
Sylvia Lin
Please go talk to your coworkers.
12 minutes ago · Like · Reply
Russ Maschmeyer
I have a relevant and impactful personal story to share about this specific topic that you can bring it to the Senate floor.
9 minutes ago · Unlike · Reply · 1
U.S. Senator Kamala Harris
Thank you for your feedback. We will message you directly to follow up.
4 minutes ago · Like · Reply
Write a reply...
Write a comment...
Post
Constituent Badges
Representatives want to hear from the people who live in their district. Show your constituent badge when you comment on your representatives' posts.
Learn More
Next
Verizon 17:54
Cancel
Find Your District
Ex: 1234 Main St, Yourtown
Enter Your Address
Enter your home address to verify who your elected representatives are. Your badge will show up when you comment on their posts.
Your address won't be displayed to others on Facebook.


Carrier 1:20 PM 100%
Search
Like
Comment
Share
Kim Ho and 147 others
Sylvia Lin
Please go talk to your coworkers.
12 minutes ago · Like · Reply
Russ Maschmeyer
I have a relevant and impactful personal story to share about this specific topic that you can bring it to the Senate floor.
9 minutes ago · Unlike · Reply · 1
U.S. Senator Kamala Harris
Thank you for your feedback. We will message you directly to follow up.
4 minutes ago · Like · Reply
Write a reply...
Write a comment...
Post
Constituent Badges
Representatives want to hear from the people who live in their district. Show your constituent badge when you comment on your representatives' posts.
Learn More
Next
Verizon 17:54
Cancel
Find Your District
Ex: 1234 Main St, Yourtown
Enter Your Address
Enter your home address to verify who your elected representatives are. Your badge will show up when you comment on their posts.
Your address won't be displayed to others on Facebook.
Your Constituent Badge is On
Your representatives will see it when you comment on their posts.
Sarah Pease
You're commenting as a constituent

Randomization at User Level

Exposure: User sees a (visible or invisible) badged comment or badging opportunity in Town Hall

Treatment (99%): Can badge, can see badges

- Not Enabled
- Enabled Badge
  - Not Constituent
  - Is Constituent
    - No Comment
    - Commented

Control (1%): Cannot badge, cannot see badges

- Is Constituent
  - Commented
  - No Comment
- Not Constituent

Randomization at User Level
Exposure: User sees a (visible or invisible) badged comment or badging opportunity in Town Hall

Treatment (99%): Can badge, can see badges

Not Enabled

Enabled Badge

Not Constituent

Is Constituent

No Comment

Commented

Control (1%): Cannot badge, cannot see badges

Is Constituent

Not Constituent

Commented

No Comment

Focus on Intent to Treat Effects

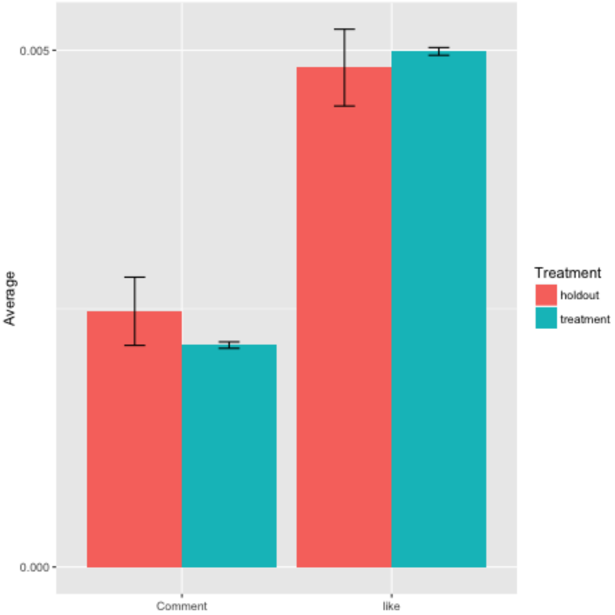
Who Enables the Badge?

Among Users who Interact with Politicians

	% Female	Age	Facebook Friends
Badged	54.3	53.9	395.0
Not Badged	46.5	49.1	475.5

Aligns with survey-based evidence in Bode (2016)

Badging Does Not Increase Politician Replies



No Heterogeneity By Office

ITT Effects By Office

Office	Like	Comment
Overall	0.0002 [-0.0002, 0.0005]	-0.0003 [-0.0007, 0.0000]
Mayor	0.003 [0.0031,0.0035]	-0.000 [-0.000, 0.000]
State Lower	-0.0001 [-0.001, 0.0008]	-0.0003 [-0.001, 0.0005]
State Upper	**0.0052** [0.004, 0.006]	**-0.006** [-0.007, -0.005]
Governor	**-0.0008** [-0.0003,-0.0001]	**-0.0002** [-0.0009, -0.0006]
House	0.0001 [-0.0016, 0.0013]	-0.0014 [-0.0001, 0.0003]
Senate	0.0001 [0.0001,0.0001]	0.0000 [0.0000, 0.0000]

Why?

Why?

- Bad product, elected officials confused

Why?

- Bad product, elected officials confused
- Constituents are not actually important

Why?

- Bad product, elected officials confused
- Constituents are not actually important
- Information matters, not on site

Elected Officials Write Longer Posts When They Do Respond

- Depart from Experimental Design
- Elected officials write longer posts $\rightsquigarrow$ more effort
- Examine how much longer responses are to badged comments

Variables	Count			Log(Count + 1)		
Badged	5.96	4.4	2.2	0.24	0.20	0.08
	(0.57)	(0.64)	(0.67)	(0.02)	(0.02)	(0.02)
Comment Length	-	0.2	0.14	-	0.01	0.01
	-	(0.01)	(0.01)	-	(0.00)	(0.00)
USA Location	-	-0.03	1.84	-	-0.03	0.08
	-	(0.48)	(3.3)	-	(0.01)	(0.09)
Text Topics	No	Yes	Yes	No	Yes	Yes
Positive	No	Yes	Yes	No	Yes	Yes
Negative	No	Yes	Yes	No	Yes	Yes
Political Words	No.	Yes	Yes	No	Yes	Yes
Politician Fixed Effects	No	No	Yes	No	No	Yes

Wrap Up

- Text as Data: Discovery, Measurement, and Causal Inference
- Lots or ongoing research in this area!
- Applications to many non-text settings
- Intersects with many other areas