Progressive Media Impact: Metrics, Evaluation, and Improvement

Gary King

Institute for Quantitative Social Science
Harvard University

(talk at The Media Consortium, San Francisco, 10/13/2011)
Goals

- Measure public discourse
- Estimate the causal effect of the progressive media on discourse
- Learn together how to make the progressive media more effective
- Take advantage of recent dramatic scientific advances:
  - Informative data now available on the public debate
  - Measurement methods: meaning, not word counts
  - Experimental designs: avoiding confounding factors
  - Causal effect estimation: less bias & inefficiency
  - New ways of working together to improve impact
Goals

- Measure public discourse
Goals

- Measure public discourse
- Estimate the causal effect of the progressive media on discourse
Goals

- **Measure** public discourse
- **Estimate the causal effect** of the progressive media on discourse
- **Learn together** how to make the progressive media more effective
Goals

- Measure public discourse
- Estimate the causal effect of the progressive media on discourse
- Learn together how to make the progressive media more effective
- Take advantage of recent dramatic scientific advances:
Goals

- **Measure** public discourse
- Estimate the *causal effect* of the progressive media on discourse
- **Learn together** how to make the progressive media more effective
- **Take advantage of** recent dramatic scientific advances:
  - Informative data now available on the public debate
Goals

- Measure public discourse
- Estimate the causal effect of the progressive media on discourse
- Learn together how to make the progressive media more effective
- Take advantage of recent dramatic scientific advances:
  - Informative data now available on the public debate
  - Measurement methods: meaning, not word counts
Goals

- **Measure** public discourse
- Estimate the **causal effect** of the progressive media on discourse
- **Learn together** how to make the progressive media more effective
- Take advantage of **recent dramatic scientific advances**:
  - Informative data now available on the public debate
  - Measurement methods: meaning, not word counts
  - Experimental designs: avoiding confounding factors
Goals

- Measure public discourse
- Estimate the causal effect of the progressive media on discourse
- Learn together how to make the progressive media more effective
- Take advantage of recent dramatic scientific advances:
  - Informative data now available on the public debate
  - Measurement methods: meaning, not word counts
  - Experimental designs: avoiding confounding factors
  - Causal effect estimation: less bias & inefficiency
Goals

- **Measure** public discourse
- Estimate the **causal effect** of the progressive media on discourse
- **Learn together** how to make the progressive media more effective
- **Take advantage of recent dramatic scientific advances:**
  - Informative data now available on the public debate
  - Measurement methods: meaning, not word counts
  - Experimental designs: avoiding confounding factors
  - Causal effect estimation: less bias & inefficiency
  - New ways of working together to improve impact
Who Frames the Public Debate?

On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president.” The next day — which the president spent in hiding — 90% approved. Was this massive opinion change, or did the 9/11 frame change how we viewed the question?

The frame: imposed by events, not the media

Public opinion polls: measuring what?

Support for Ban on “Partial Birth” Abortion

Supporters: use “baby”

Opponents: use “fetus”

In surveys, “baby” increases support for the ban by ≈30%!

Control the frame & you control the debate and policy outcome

Gary King (Harvard)
On 9/10/2001, 55% of Americans approved of the way George W. Bush was "handling his job as president." The next day — which the president spent in hiding — 90% approved. Was this massive opinion change, or did the 9/11 frame change how we viewed the question? The frame: imposed by events, not the media.
“Do you approve of how George W. Bush is handling his job?”

On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.

Support for Ban on “Partial Birth” Abortion

Supporters: use “baby”

Opponents: use “fetus”

In surveys, “baby” increases support for the ban by ≈ 30%!
Who Frames the Public Debate?

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.

Support for Ban on “Partial Birth” Abortion

- Supporters: use “baby”
- Opponents: use “fetus”

In surveys, “baby” increases support for the ban by $\approx 30\%$!
“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
Who Frames the Public Debate?

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
- The frame: imposed by events, not the media
“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
- The frame: imposed by events, not the media
- Public opinion polls: measuring what?
Who Frames the Public Debate?

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
- The frame: imposed by events, not the media
- Public opinion polls: measuring what?

Support for Ban on “Partial Birth” Abortion
Who Frames the Public Debate?

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
- The frame: imposed by events, not the media
- Public opinion polls: measuring what?

Support for Ban on “Partial Birth” Abortion

- Supporters: use “baby”
Who Frames the Public Debate?

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?

The frame: imposed by events, not the media

Public opinion polls: measuring what?

Support for Ban on “Partial Birth” Abortion

- Supporters: use “baby”
- Opponents: use “fetus”
Who Frames the Public Debate?

“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
- The frame: imposed by events, not the media
- Public opinion polls: measuring what?

Support for Ban on “Partial Birth” Abortion

- Supporters: use “baby”
- Opponents: use “fetus”
- In surveys, “baby” increases support for the ban by ≈ 30%!
“Do you approve of how George W. Bush is handling his job?”

- On 9/10/2001, 55% of Americans approved of the way George W. Bush was “handling his job as president”.
- The next day — which the president spent in hiding — 90% approved.
- Was this massive opinion change, or did the 9/11 frame change how we viewed the question?
- The frame: imposed by events, not the media
- Public opinion polls: measuring what?

Support for Ban on “Partial Birth” Abortion

- Supporters: use “baby”
- Opponents: use “fetus”
- In surveys, “baby” increases support for the ban by \( \approx 30\% \)
- Control the frame & you control the debate and policy outcome
Public opinion regarding PBA is measured infrequently, making it impossible to include aggregate opinion in the time series analysis. Instead, responses to 12 similarly worded questions from 1996 to 2000 were charted (each asked respondents to indicate their level of support for a PBA ban). Figure 1 shows that aggregate support for a ban seemed to rise and fall in tandem with the proportion of "baby" usage in news stories about PBA.

Figure 1: Relationship between media discourse and public support for partial-birth abortion ban.

Note: The media series represents a 5-week moving average of the proportion of "baby" mentions. Public opinion data come from similarly worded questions in surveys by Gallup (squares) and Princeton Survey Research Associates (triangles). These items ask respondents about their level of support for a ban on partial-birth abortion. Question wording can be obtained from LexisNexis or the iPoll database at the Roper Center for Public Opinion.

Table 3: Granger Causality Tests Examining the Elite–News Relationship

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Coefficient Block</th>
<th>Block x^2</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>Congressional rhetoric</td>
<td>11.13</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>22.24</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Editorials</td>
<td>2.24</td>
<td>.53</td>
</tr>
<tr>
<td>Congressional rhetoric</td>
<td>News</td>
<td>13.22</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Editorials</td>
<td>4.31</td>
<td>.23</td>
</tr>
<tr>
<td>Editorials</td>
<td>Congressional rhetoric</td>
<td>7.41</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>10.03</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Editorials</td>
<td>5.14</td>
<td>.08</td>
</tr>
</tbody>
</table>

Note: Arrows indicate Granger causality from coefficient block to the dependent variable. Three lags of each independent variable are included in the model. Diagnostic tests indicate no evidence of autocorrelation (up to four lags). Model residuals are white noise. Analyses were done using STATA 8.2.
“Partial Birth” Abortion Ban: Who controls the frame?

Public opinion regarding PBA is measured infrequently, making it impossible to include aggregate opinion in the time series analysis. Instead, responses to 12 similarly worded questions from 1996 to 2000 were charted (each asked respondents to indicate their level of support for a PBA ban). Figure 1 shows that aggregate support for a ban seemed to rise and fall in tandem with the proportion of “baby” usage in news stories about PBA.

Figure 1: Relationship between media discourse and public support for partial-birth abortion ban.

Note: The media series represents a 5-week moving average of the proportion of “baby” mentions. Public opinion data come from similarly worded questions in surveys by Gallup (squares) and Princeton Survey Research Associates (triangles). These items ask respondents about their level of support for a ban on partial-birth abortion. Question wording can be obtained from LexisNexis or the iPoll database at the Roper Center for Public Opinion.

Table 3: Granger Causality Tests Examining the Elite–News Relationship

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Coefficient Block x</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>Congress rhetoric</td>
<td>11.13</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>22.24</td>
</tr>
<tr>
<td></td>
<td>Editorials</td>
<td>2.24</td>
</tr>
<tr>
<td>Congressional rhetoric</td>
<td>News</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>Editorials</td>
<td>4.31</td>
</tr>
<tr>
<td>Editorials</td>
<td>Congressional rhetoric</td>
<td>7.41</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>10.03</td>
</tr>
<tr>
<td></td>
<td>Editorials</td>
<td>5.14</td>
</tr>
</tbody>
</table>

Note: Arrows indicate Granger causality from coefficient block to the dependent variable. Three lags of each independent variable are included in the model. Diagnostic tests indicate no evidence of autocorrelation (up to four lags). Model residuals are white noise. N = 386.

Analyses were done using STATA 8.2.

A. F. Simon & J. Jerit
ª 2007 International Communication Association
“Partial Birth” Abortion Ban: Who controls the frame?

When “baby” is used in the media, support for PBA ban increases
Partial Birth Abortion Ban: Who controls the frame?

Public opinion regarding PBA is measured infrequently, making it impossible to include aggregate opinion in the time series analysis. Instead, responses to 12 similarly worded questions from 1996 to 2000 were charted (each asked respondents to indicate their level of support for a PBA ban). Figure 1 shows that aggregate support for a ban seemed to rise and fall in tandem with the proportion of “baby” usage in news stories about PBA.

![Graph showing relationship between media discourse and public support for partial-birth abortion ban.](image)

- When “baby” is used in the media, support for PBA ban increases.
- Did the media influence public support or did public support influence the media?
Partial Birth Abortion Ban: Who controls the frame?

Public opinion regarding PBA is measured infrequently, making it impossible to include aggregate opinion in the time series analysis. Instead, responses to 12 similarly worded questions from 1996 to 2000 were charted (each asked respondents to indicate their level of support for a PBA ban). Figure 1 shows that aggregate support for a ban seemed to rise and fall in tandem with the proportion of “baby” usage in news stories about PBA.

When “baby” is used in the media, support for PBA ban increases

Did the media influence public support or did public support influence the media? How can we tell?
When “baby” is used in the media, support for PBA ban increases.

Did the media influence public support or did public support influence the media? How can we tell?

Do “baby” and “fetus” word counts even measure what we intend?
The Plan, 1: Collect New Data

Data on Public Discourse

Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent.

Now:

Hallway conversations appear in the 1.4B social media posts a week

Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.

Data on Media Outlets

Work together to find or create

Existing data on outputs (web traffic, donations, comments, emails, phone logs, etc.)

Existing data on content: e.g., tagged story databases, web artifacts, listener information, etc.

New systematic data from texts of stories, audio-to-text, etc.
The Plan, 1: Collect New Data

Data on Public Discourse

Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent.

Now:

- Hallway conversations appear in the 1.4B social media posts a week.
- Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.

Data on Media Outlets

Work together to find or create existing data on outputs (web traffic, donations, comments, emails, phone logs, etc.).

Existing data on content: e.g., tagged story databases, web artifacts, listener information, etc.

New systematic data from texts of stories, audio-to-text, etc.

Gary King (Harvard)
Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent.
The Plan, 1: Collect New Data

Data on Public Discourse

- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent
- Now:
The Plan, 1: Collect New Data

Data on Public Discourse

- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent
- Now:
  - Hallway conversations appear in the 1.4B social media posts a week
The Plan, 1: Collect New Data

Data on Public Discourse

- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent

- Now:
  - Hallway conversations appear in the 1.4B social media posts a week
  - Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.
The Plan, 1: Collect New Data

Data on Public Discourse

- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent

- Now:
  - Hallway conversations appear in the 1.4B social media posts a week
  - Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.

Data on Media Outlets

Work together to find or create
The Plan, 1: Collect New Data

Data on Public Discourse
- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent
- Now:
  - Hallway conversations appear in the 1.4B social media posts a week
  - Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.

Data on Media Outlets
Work together to find or create
- Existing data on outputs (web traffic, donations, comments, emails, phone logs, etc.)
The Plan, 1: Collect New Data

Data on Public Discourse

- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent
- Now:
  - Hallway conversations appear in the 1.4B social media posts a week
  - Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.

Data on Media Outlets

Work together to find or create

- Existing data on outputs (web traffic, donations, comments, emails, phone logs, etc.)
- Existing data on content: e.g., tagged story databases, web artifacts, listener information, etc.
The Plan, 1: Collect New Data

**Data on Public Discourse**
- Until recently, measures of public discourse were inadequate: public opinion polls, content analyses of newspaper editorials; systematic, real time, informative data nonexistent
- **Now:**
  - Hallway conversations appear in the 1.4B social media posts a week
  - Growing sources: blogs, Twitter, Facebook, forums, chat rooms, etc.

**Data on Media Outlets**
Work together to find or create
- **Existing data on outputs** (web traffic, donations, comments, emails, phone logs, etc.)
- **Existing data on content:** e.g., tagged story databases, web artifacts, listener information, etc.
- **New systematic data** from texts of stories, audio-to-text, etc.
The Plan, 2: Adapt New Methods to Understand Billions of Social Media Posts, & News Stories

Until recently, methods were inadequate:
- Misleading word counts
- Impossible amounts of reading

Now, new statistical methods for text analytics:
- Accurately summarize huge volumes of information
- Better than humans alone, or computers alone, we amplify human intelligence
- Works fast, in near real time
- Can measure media frames (about politicians or policies) by volume, sentiment, topics, perspectives — or any categories we think of
- Measure as far back in time as feasible (6-24 months?)
- Design systematic measurement going forward
Until recently, methods were inadequate:
The Plan, 2: Adapt New Methods to Understand Billions of Social Media Posts, & News Stories

- Until recently, methods were inadequate:
  - Misleading word counts
The Plan, 2: Adapt New Methods to Understand Billions of Social Media Posts, & News Stories

- Until recently, methods were inadequate:
  - Misleading word counts
  - Impossible amounts of reading

Now, new statistical methods for text analytics:
- Accurately summarize huge volumes of information
- Better than humans alone, or computers alone, we amplify human intelligence
- Works fast, in near real time
- Can measure media frames (about politicians or policies) by volume, sentiment, topics, perspectives — or any categories we think of
- Measure as far back in time as feasible (6-24 months?)
- Design systematic measurement going forward
The Plan, 2: Adapt New Methods to Understand Billions of Social Media Posts, & News Stories

- Until recently, methods were inadequate:
  - Misleading word counts
  - Impossible amounts of reading
- Now, new statistical methods for text analytics:
Until recently, methods were inadequate:

- Misleading word counts
- Impossible amounts of reading

Now, new statistical methods for text analytics:

- Accurately summarize huge volumes of information
Until recently, methods were inadequate:
  - Misleading word counts
  - Impossible amounts of reading

Now, new statistical methods for text analytics:
  - Accurately summarize huge volumes of information
  - Better than humans alone, or computers alone, we amplify human intelligence
Until recently, methods were inadequate:

- Misleading word counts
- Impossible amounts of reading

Now, new statistical methods for text analytics:

- Accurately summarize huge volumes of information
- Better than humans alone, or computers alone, we amplify human intelligence
- Works fast, in near real time
The Plan, 2: Adapt New Methods to Understand Billions of Social Media Posts, & News Stories

- Until recently, methods were inadequate:
  - Misleading word counts
  - Impossible amounts of reading
- Now, new statistical methods for text analytics:
  - Accurately summarize huge volumes of information
  - Better than humans alone, or computers alone, we amplify human intelligence
  - Works fast, in near real time
  - Can measure media frames (about politicians or policies) by volume, sentiment, topics, perspectives — or any categories we think of
The Plan, 2: Adapt New Methods to Understand Billions of Social Media Posts, & News Stories

- Until recently, methods were inadequate:
  - Misleading word counts
  - Impossible amounts of reading
- Now, new statistical methods for text analytics:
  - Accurately summarize huge volumes of information
  - Better than humans alone, or computers alone, we amplify human intelligence
  - Works fast, in near real time
  - Can measure media frames (about politicians or policies) by volume, sentiment, topics, perspectives — or any categories we think of
  - Measure as far back in time as feasible (6-24 months?)
Until recently, methods were inadequate:
- Misleading word counts
- Impossible amounts of reading

Now, new statistical methods for text analytics:
- Accurately summarize huge volumes of information
- Better than humans alone, or computers alone, we *amplify human intelligence*
- Works fast, in near real time
- Can measure media frames (about politicians or policies) by volume, sentiment, topics, perspectives — or any categories we think of
- Measure as far back in time as feasible (6-24 months?)
- Design systematic measurement going forward
How can humans understand large numbers of social media posts (or news stories)?

Read & interpret: infeasible

Sort into a few categories; track category percentages over time

Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor.

How to sort billions of social media posts into categories?

Classify by hand: infeasible for more than a sample at one point in time

Guess what words imply which categories: inaccurate

Use "machine learning" methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)

Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate

Finding the needle in the haystack (Google search) \( \neq \) characterizing the haystack, and so new methods were required

Gary King (Harvard)
Progressive Media Impact
How can humans understand large numbers of social media posts (or news stories)?

- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor.

- How to sort billions of social media posts into categories?
  - Classify by hand: infeasible for more than a sample at one point in time
  - Guess what words imply which categories: inaccurate
  - Use "machine learning" methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)
  - Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate

Finding the needle in the haystack (Google search) ≠ characterizing the haystack, and so new methods were required.

Gary King (Harvard)
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is

Example categories:
1. A responsible effort to fix budget deficit
2. An attack on public sector workers
3. An attack on the progressive movement generally
4. Due to an arrogant publicity-hungry Governor
How can humans understand large numbers of social media posts (or news stories)?
- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit

How to sort billions of social media posts into categories?
- Classify by hand: infeasible for more than a sample at one point in time
- Guess what words imply which categories: inaccurate
- Use "machine learning" methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)
- Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate

Finding the needle in the haystack (Google search) ≠ characterizing the haystack, and so new methods were required
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally

Classification by hand: infeasible for more than a sample at one point in time

Guess what words imply which categories: inaccurate

Use “machine learning” methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)

Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor

How to sort billions of social media posts into categories?

Classify by hand: infeasible for more than a sample at one point in time

Guess what words imply which categories: inaccurate

Use “machine learning” methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)

Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor

How to sort billions of social media posts into categories?

- Classify by hand: infeasible for more than a sample at one point in time
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor

How to sort billions of social media posts into categories?

- Classify by hand: infeasible for more than a sample at one point in time
- Guess what words imply which categories: inaccurate
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor

How to sort billions of social media posts into categories?

- Classify by hand: infeasible for more than a sample at one point in time
- Guess what words imply which categories: inaccurate
- Use “machine learning” methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)
How can humans understand large numbers of social media posts (or news stories)?

- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor

How to sort billions of social media posts into categories?

- Classify by hand: infeasible for more than a sample at one point in time
- Guess what words imply which categories: inaccurate
- Use “machine learning” methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)
- Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate
How can humans understand large numbers of social media posts (or news stories)?
- Read & interpret: infeasible
- Sort into a few categories; track category percentages over time
- Example of categories: Wisconsin budget standoff is (1) a responsible effort to fix budget deficit, (2) an attack on public sector workers, (3) an attack on the progressive movement generally, or (4) due to an arrogant publicity-hungry Governor

How to sort billions of social media posts into categories?
- Classify by hand: infeasible for more than a sample at one point in time
- Guess what words imply which categories: inaccurate
- Use “machine learning” methods with hand-coded training set to automatically classify: at best 60-70% accuracy (useful for Google searches, useless for category percentages)
- Use new statistical method with hand-coded training set to estimate category percentages (without individual classifications): Extremely accurate
- Finding the needle in the haystack (Google search) ≠ characterizing the haystack, and so new methods were required
The Plan, 3: Use New Methods of Causal Inference for Observational Data

Until recently:

- Estimating causal effects from observational data was very risky and often illusory

Now, new methods:

- Give strong hints about causal relationships
- Reveal all hidden assumptions (ignorability, interference, etc.)

We'll use new methods to:

- Estimate the effect of progressive media on use of media frames, & sentiment about them, in public discourse
- Develop strong testable hypotheses about what to do next

We'll do everything that can be done this way before turning to experiments (and using your time!)
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- **Until recently:** Estimating causal effects from observational data was very risky and often illusory.
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- **Until recently:** Estimating causal effects from observational data was very risky and often illusory
- **Now, new methods:**

- Give strong hints about causal relationships
- Reveal all hidden assumptions (ignorability, interference, etc.)
- Help inform what experiments to run
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- **Until recently:** Estimating causal effects from observational data was very risky and often illusory
- **Now, new methods:**
  - Give strong hints about causal relationships
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- **Until recently:** Estimating causal effects from observational data was very risky and often illusory
- **Now, new methods:**
  - Give strong hints about causal relationships
  - Reveal all hidden assumptions (ignorability, interference, etc.)
Until recently: Estimating causal effects from observational data was very risky and often illusory

Now, new methods:
- Give strong hints about causal relationships
- Reveal all hidden assumptions (ignorability, interference, etc.)
- Help inform what experiments to run

We'll use new methods to:
- Estimate the effect of progressive media on use of media frames, & sentiment about them, in public discourse
- Develop strong testable hypotheses about what to do next
- We'll do everything that can be done this way before turning to experiments (and using your time!)
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- Until recently: Estimating causal effects from observational data was very risky and often illusory
- Now, new methods:
  - Give strong hints about causal relationships
  - Reveal all hidden assumptions (ignorability, interference, etc.)
  - Help inform what experiments to run
- We’ll use new methods to:
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- **Until recently:** Estimating causal effects from observational data was very risky and often illusory
- **Now, new methods:**
  - Give strong hints about causal relationships
  - Reveal all hidden assumptions (ignorability, interference, etc.)
  - Help inform what experiments to run
- **We’ll use new methods to:**
  - Estimate the effect of progressive media on use of media frames, & sentiment about them, in public discourse
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- **Until recently:** Estimating causal effects from observational data was very risky and often illusory
- **Now, new methods:**
  - Give strong hints about causal relationships
  - Reveal all hidden assumptions (ignorability, interference, etc.)
  - Help inform what experiments to run
- **We’ll use new methods to:**
  - Estimate the effect of progressive media on use of media frames, & sentiment about them, in public discourse
  - Develop strong testable hypotheses about what do next
The Plan, 3: Use New Methods of Causal Inference for Observational Data

- Until recently: Estimating causal effects from observational data was very risky and often illusory.
- Now, new methods:
  - Give strong hints about causal relationships
  - Reveal all hidden assumptions (ignorability, interference, etc.)
  - Help inform what experiments to run
- We’ll use new methods to:
  - Estimate the effect of progressive media on use of media frames, & sentiment about them, in public discourse
  - Develop strong testable hypotheses about what do next
  - We’ll do everything that can be done this way before turning to experiments (and using your time!)
Observational analysis: some evidence of what might have worked in the past

Experiments: gold standard for evaluating causal claims
Experiments benefits from: investigator control (usually randomized) over the "treatment" (drugs, stories)
What randomized treatments get us: a variable unrelated to all confounders

How experiments fail: Insensitivity to research subjects, political interests, or local context
Examples: Drug trials; Mexico's Anti-Poverty program
⇝ solutions always exist; we'll find them together

New experimental designs: let randomization survive disruption
Observational analysis: some evidence of what might have worked in the past
The Plan, 4: Use New Experimental Designs

- **Observational analysis:** some evidence of what might have worked in the past
- **Experiments:** gold standard for evaluating causal claims
The Plan, 4: Use New Experimental Designs

- **Observational analysis**: some evidence of what might have worked in the past
- **Experiments**: gold standard for evaluating causal claims
- **Experiments benefits from**: investigator control (usually randomized) over the “treatment” (drugs, stories)
Observational analysis: some evidence of what might have worked in the past

Experiments: gold standard for evaluating causal claims

Experiments benefits from: investigator control (usually randomized) over the “treatment” (drugs, stories)

What randomized treatments get us: a variable unrelated to all confounders
The Plan, 4: Use New Experimental Designs

- **Observational analysis**: some evidence of what might have worked in the past
- **Experiments**: gold standard for evaluating causal claims
- **Experiments benefits from**: investigator control (usually randomized) over the “treatment” (drugs, stories)
- **What randomized treatments get us**: a variable unrelated to all confounders
- **How experiments fail**: 

...
Observational analysis: some evidence of what might have worked in the past

Experiments: gold standard for evaluating causal claims

Experiments benefits from: investigator control (usually randomized) over the “treatment” (drugs, stories)

What randomized treatments get us: a variable unrelated to all confounders

How experiments fail:
  - Insensitivity to research subjects, political interests, or local context
Observational analysis: some evidence of what might have worked in the past

Experiments: gold standard for evaluating causal claims

Experiments benefits from: investigator control (usually randomized) over the “treatment” (drugs, stories)

What randomized treatments get us: a variable unrelated to all confounders

How experiments fail:
- Insensitivity to research subjects, political interests, or local context
- Examples: Drug trials; Mexico’s Anti-Poverty program
The Plan, 4: Use New Experimental Designs

- **Observational analysis:** some evidence of what might have worked in the past
- **Experiments:** gold standard for evaluating causal claims
- **Experiments benefits from:** investigator control (usually randomized) over the “treatment” (drugs, stories)
- **What randomized treatments get us:** a variable unrelated to all confounders
- **How experiments fail:**
  - Insensitivity to research subjects, political interests, or local context
  - Examples: Drug trials; Mexico’s Anti-Poverty program
  - Solutions always exist; we’ll find them together
Observational analysis: some evidence of what might have worked in the past

Experiments: gold standard for evaluating causal claims

Experiments benefits from: investigator control (usually randomized) over the “treatment” (drugs, stories)

What randomized treatments get us: a variable unrelated to all confounders

How experiments fail:
  - Insensitivity to research subjects, political interests, or local context
  - Examples: Drug trials; Mexico’s Anti-Poverty program
  - solutions always exist; we’ll find them together

New experimental designs: let randomization survive disruption
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we're indifferent to being above or below the fold; the order of stories to present)

Example: Mexico's Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will start small (probably individual media outlets), be sequential (each will build on the previous), and grow in extent (larger experiments, with more outlets, as we learn more)

Our goal: Turn evidence of what has worked into increasingly effective strategies
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will start small (probably individual media outlets), be sequential (each will build on the previous), and grow in extent (larger experiments, with more outlets, as we learn more)

Our goal: Turn evidence of what has worked into increasingly effective strategies
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)

- Find otherwise arbitrary decisions to randomize (e.g., a story we're indifferent to being above or below the fold; the order of stories to present)

Example: Mexico's Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will start small (probably individual media outlets), be sequential (each will build on the previous), and grow in extent (larger experiments, with more outlets, as we learn more)

Our goal: Turn evidence of what has worked into increasingly effective strategies
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarcest resource: your time (so observational work first)
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarest resource: your time (so observational work first)

Experiments will start small (probably individual media outlets)
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will start small (probably individual media outlets), be sequential (each will build on the previous)
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will **start small** (probably individual media outlets), be **sequential** (each will build on the previous), and **grow in extent** (larger experiments, with more outlets, as we learn more)
Examples of how to experiment without infringing on editorial discretion:

- Offer optional small grants to media outlets to encourage covering particular stories (everyone will get a chance, no requirements, all decision making over content remains with organization)
- Find otherwise arbitrary decisions to randomize (e.g., a story we’re indifferent to being above or below the fold; the order of stories to present)
- Example: Mexico’s Health Insurance evaluation

Our scarcest resource: your time (so observational work first)

Experiments will start small (probably individual media outlets), be sequential (each will build on the previous), and grow in extent (larger experiments, with more outlets, as we learn more)

Our goal: Turn evidence of what has worked into increasingly effective strategies
For More Information

GKing.Harvard.edu