Measuring the Persuasive Power of Political Arguments

PSA Political Methodology Group Annual Conference

Jack Blumenau and Benjamin Lauderdale

10th January, 2020

Which types of political rhetoric are most effective?

- *Descriptively* we know that politicians invest much time and effort in crafting rhetorical appeals, and these often share common types.

- It is *normatively* important to establish which forms of rhetorical appeal are successful.

- Despite growing empirical work in this area, few papers try to quantify the *relative* power of different argument types.

Our contribution: an experimental design and modelling framework to measure the relative effectiveness of different types of political rhetoric.

Existing approaches (a caricature)

A common structure for papers in political communication is that researchers have a **substantive interest** in the efficacy of latent concept:

"Does rhetoric of type X affect public opinion/voter behaviour?"

This is matched with an **experimental design** which randomizes few/single texts that implement the concept:

Treatment text featuring rhetoric of type X versus control without type X

Most commonly, these texts are relevant to **single policy areas**, but conclusions drawn are more general:

"Rhetoric of type X is an important cause of behaviour Y."

Existing approaches (challanges)

1. **Few policy areas**

    - → single-context studies are not informative about whether effects *generalise* to a broader population of political issues

2. **Few texts per latent treatment**

    - → any given implementation of a latent treatment may be *confounded* by other features of the text

3. **Few comparisons of relative effectiveness of different elements**

    - → determining whether some types of rhetoric are more effective than others can have significant *normative implications*

Our experimental design & modelling strategy addresses these challenges.

Experimental design

Policies, elements, and arguments

Three basic units that define our experimental design

1. **Policy issues**

   - An issue subject to political debate, where the government could take action
   - We include 12 policies

2. **Rhetorical elements**

   - A feature of political argument that emphasises the desirability or otherwise of a given policy
   - We include 14 elements

3. **Arguments**

   - A text that makes the case either in favour or against a policy.
   - In our design, each argument employs a single rhetorical element
   - We include 12 x 14 x 2 = 336 arguments

Policy issues

Building a third runway at Heathrow
Closing large retail stores on Boxing Day
Extending the Right to Buy
Extension of surveillance powers in the UK
Fracking in the UK
Nationalisation of the railways in the UK
Quotas for women on corporate boards
Reducing the legal restrictions on cannabis use
Reducing university tuition fees
Renewing Trident
Spending 0.7% of GDP on overseas aid
Sugar tax in the UK

Elements

Appeal to fairness	Appeal to expertise
Costs vs benefits arguments	Appeal to history
Country comparison	Appeal to national greatness
Crisis	Appeal to populism
Side-effects	Common sense
Metaphor	Morality
Ad hominem	Public opinion

Elements

Appeal to fairness	Appeal to expertise
Costs vs benefits arguments	Appeal to history
Country comparison	Appeal to national greatness
Crisis	Appeal to populism
Side-effects	Common sense
Metaphor	Morality
Ad hominem	Public opinion

E.g. Boudreau and MacKenzie, (2014); Dewan, Humphreys and Rubenson, (2014); Atkins and Finlayson, (2013)

Elements

Appeal to fairness	Appeal to expertise
Costs vs benefits arguments	Appeal to history
Country comparison	Appeal to national greatness
Crisis	Appeal to populism
Side-effects	Common sense
Metaphor	Morality
Ad hominem	Public opinion

E.g. Bos et al (2013); Hameleers et al (2017); Hameleers and Schmuck, (2017)

Elements

Appeal to fairness	Appeal to expertise
Costs vs benefits arguments	Appeal to history
Country comparison	Appeal to national greatness
Crisis	Appeal to populism
Side-effects	Common sense
Metaphor	Morality
Ad hominem	Public opinion

E.g. Thibodeau and Boroditsky (2011); Schlesinger and Lau (2000); Bougher (2012); Charteris-Black (2011)

Arguments

Morality argument in favour of Quotas for women on corporate boards

> *It is a moral failure, in a society that aspires to be equal, to have such extraordinary low numbers of women in important positions.*

Arguments

Morality argument against the Extension of surveillance powers

It is immoral to engage in the mass collection of private data from people who have not been accused of doing anything wrong.

Randomization scheme

For each respondent, randomly select

1. 1 of 12 policy issues

2. whether arguments will be on same or different sides of the issue

3. 2 arguments subject to the above

4 responses from 3317 respondents = 13268 observations.

Prompt

YouGov

Building a third runway at Heathrow

London's Heathrow airport has two runways that are currently operating at full capacity. Some people are in favour of building a third runway at Heathrow ("for"), others are opposed ("against").

Please read the following **arguments for and against** building a third Runway at Heathrow.

Argument One (For)	Argument Two (Against)
The Airports Commission, an independent body established to study the issue, have argued that expanding Heathrow is "the most effective option to address the UK's aviation capacity challenge".	Building a third runway would be giving a blank cheque to the foreign-owned multinational company that runs Heathrow.

Which of these arguments do you find more persuasive?

| Argument **one** is more persuasive | They are about the same | Argument **two** is more persuasive |

>

Responses

Responses to our experiment provide information on the relative persuasiveness of different arguments:

$$Y_i \in \left\{ \begin{array}{l} 1 = \text{Argument 2 is more persuasive} \\ 2 = \text{They are about the same} \\ 3 = \text{Argument 1 is more persuasive} \end{array} \right.$$

Respondents gave the "They are about the same" response in 32% of same-side comparisons, and 25% of different-side comparisons.

Modelling persuasiveness

First-level model

We have $J = 336$ arguments, denoted $j = 1, ..., J$ which we present to respondents in paired comparisons, and which we model via:

$$log \left[\frac{P(Y_{j,j'} \leq k)}{P(Y_{j,j'} > k)} \right] = \theta_k + \alpha_j - \alpha'_j$$

- Each argument j is described by a single "strength" parameter α_j.
- θ_k is the cutpoint for response category k

First-level model

We have $J = 336$ arguments, denoted $j = 1, ..., J$ which we present to respondents in paired comparisons, and which we model via:

$$log \left[\frac{P(Y_{j,j'} \leq k)}{P(Y_{j,j'} > k)} \right] = \theta_k + \alpha_j - \alpha'_j$$

Intuition

α_j will be larger (and argument j more powerful) when
1. argument j "wins" more pairwise contests
2. the arguments defeated by j are themselves powerful

Second-level model

We model the persuasive power of individual arguments using a hierarchical model where we have parameters to capture
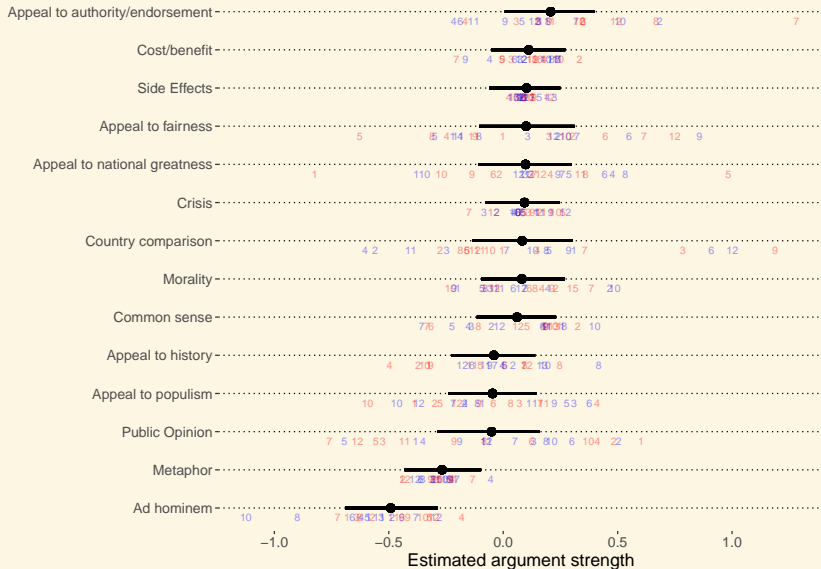
1. The average persuasiveness of the rhetorical elements

2. The variance of arguments within each element

3. The baseline persuasiveness of being on the "for" or "against" side of an issue and allows us to distinguish persuasive arguments from popular policies

Results

Rhetorical element effects
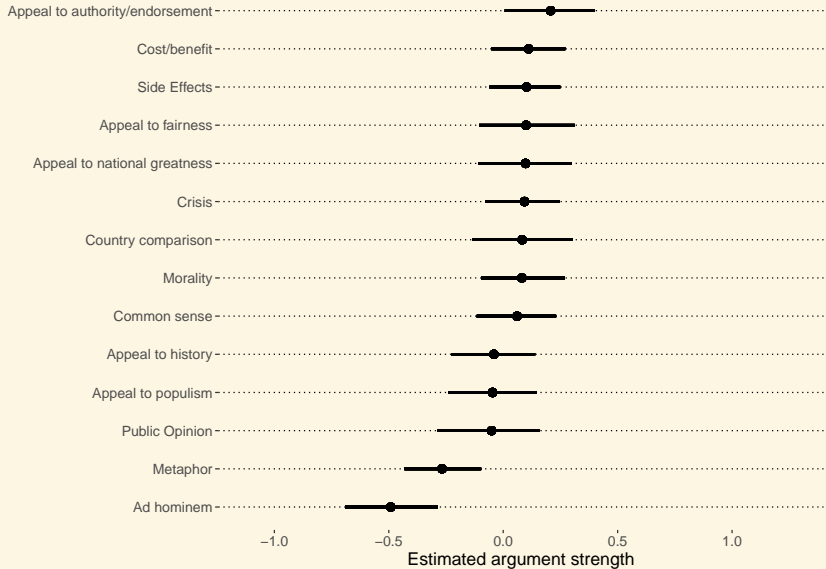
Rhetorical element effects

Argument effects

Argument effects

Strongest argument (appeal to expertise, against quotas for women):

The CEO of Burberry summed up the argument against quotas well recently by arguing 'Just put the best person into the job. It is not about gender, it is about experience, leadership and vision'.

Argument effects

Strongest argument (appeal to expertise, against quotas for women):

The CEO of Burberry summed up the argument against quotas well recently by arguing 'Just put the best person into the job. It is not about gender, it is about experience, leadership and vision'.

Weakest argument (ad hominem, in favour of renewing Trident):

Only hippies and cowards think it is a good idea to unilaterally drop our nuclear deterrant.

Sources of heterogeneity

Why do we observe different effects across arguments?

Why do we observe different effects across arguments?

1. **Issue-level heterogeneity**

   - Effects of an element may be larger for some policy issues than others
   - Implications mostly relate to external validity

Sources of heterogeneity

Why do we observe different effects across arguments?

1. **Issue-level heterogeneity**
   - Effects of an element may be larger for some policy issues than others
   - Implications mostly relate to external validity

2. **Implementation-level heterogeneity**
   - We might be bad/good at writing some types of arguments
   - This is a more pervasive problem when treating experimental texts as canonical examples of a latent concept

Why do we observe different effects across arguments?

1. **Issue-level heterogeneity**

   - Effects of an element may be larger for some policy issues than others
   - Implications mostly relate to external validity

2. **Implementation-level heterogeneity**

   - We might be bad/good at writing some types of arguments
   - This is a more pervasive problem when treating experimental texts as canonical examples of a latent concept

Note that we cannot distinguish between these given our data.

Sources of heterogeneity

Why do we observe different effects across arguments?

1. **Issue-level heterogeneity**

   · Effects of an element may be larger for some policy issues than others
   · Implications mostly relate to external validity

2. **Implementation-level heterogeneity**

   · We might be bad/good at writing some types of arguments
   · This is a more pervasive problem when treating experimental texts as canonical examples of a latent concept

Note that we cannot distinguish between these given our data. Note also that you cannot distinguish between these in your experiment!

Conclusion

Limitations and caveats

1. The findings here relate to the sample of statements we wrote

   - E.g. we, or the politicians we mimic, may be bad at writing *ad hominem* arguments

2. If research questions pertain to specific policy areas, heterogeneity across issues may not be a concern

   - The same does not apply for heterogeneity across *implementations*

3. We examine "persuasiveness" and not "persuasion"

   - Respondents may not, in fact, change their policy endorsements in response to the arguments they say are relatively persuasive

Conclusion

- Methodological contributions
  - *Experimental design* which reduces reliance on single-text treatment implementations
  - *Modelling strategy* which allows us to make statements about the heterogeneity of argument effectiveness and characterise other quantities of interest via a multi-level model

- Results and implications
  - Modest differences in average persuasiveness of rhetorical elements
  - Large differences between arguments of the same type
  - Researchers interested in evaluating the effectiveness of latent concepts should test multiple implementations of those concepts

Extra slides

Extensions

Our modelling framework allows us to...

1. ...control for (observable) potentially confounding text features

   - Add vector of controls (length; tone; readability etc) to second-level
   - Very minor effects on estimates of element strength

2. ...assess whether different groups of respondents are persuaded by the same arguments in the aggregate

   - Relatively high correlation in perceptions of argument persuasiveness

3. ...assess the similarity of estimates from same-side and different-sides argument comparisons

   - Argument strength is highly correlated for same-side and different-side estimates
   - Same-side estimates are more predictive of respondent choice

Controlling for confounders

Multiple implementations per latent treatment reduces fears of argument-level confounding, but we can also directly *control* for potentially confounding features in our treatment texts.

$$\alpha_j = \delta_{p(j),s(j)} + \mu_{e(j)} + \sum_{k=1}^{K} \gamma_k x_{k,j} + \nu_j$$

$$\mu_e \sim N(0, \omega)$$

$$\nu_j \sim N(0, \sigma_{e(j)})$$

- $x_k \rightarrow$ text-level features, including length, readability, positive and negative tone, emotional language, fact-based language
- γ_k are average linear effects of text-feature k on argument strength

Controlling for confounders