The Variable Persuasiveness of Political Rhetoric [*]

Jack Blumenau   *University College London*

Benjamin E Lauderdale   *University College London*

Which types of political rhetoric are most persuasive? Politicians make arguments that share common rhetorical elements, including metaphor, ad hominem attacks, appeals to expertise, moral appeals, and many others. However, political arguments are also highly multidimensional, making it difficult to assess the relative persuasive power of these elements. We report on a novel experimental design which assesses the relative persuasiveness of a large number of arguments that deploy a set of rhetorical elements to argue for and against proposals across a range of UK political issues. We find modest differences in the average effectiveness of rhetorical elements shared by many arguments, but also large variation in the persuasiveness of arguments of the same rhetorical type across issues. In addition to revealing that some argument-types are more effective than others in shaping public opinion, these results have important implications for the interpretation of survey-experimental studies in the field of political communication.

Word count: 9998

Introduction

Politicians invest time and effort in crafting arguments to present to voters, and the arguments that they make often deploy common rhetorical elements. Regardless of the specific policy at stake, politicians can draw on endorsements from relevant authorities; emphasise a moral rationale; carefully articulate costs and benefits; impugn the motives of opposition actors; present evidence from historical or other countries' experiences; and so on. While interest in rhetorical strategies has sustained over the course of millennia (Aristotle, c.322 BCE; Rhetorica ad herennium, c.80 BCE; Riker, 1990; Charteris-Black, 2011), and more recent work has begun to test the efficacy of different communication strategies (Loewen, Rubenson and Spirling, 2012; Thibodeau and Boroditsky, 2011; Schlesinger and Lau, 2000; Bougher, 2012; Lau, Sigelman and Rovner, 2007; Bos, Van Der Brug and De Vreese, 2013; Hameleers, Bos and de Vreese, 2017; Hameleers and Schmuck, 2017; Jung, Forthcoming; Nelson, 2004; Boudreau and MacKenzie, 2014; Jerit, 2009), making general statements about the *relative* performance of particular rhetorical strategies is difficult because arguments are so highly multidimensional. Arguments deploy common elements, but they also vary in many other ways that might make certain strategies more effective in some implementations than others. As a result, empirical research has rarely moved beyond demonstrating non-zero effectiveness of specific types of arguments that politicians employ in particular domains. As a consequence, "scholars still understand little about the factors that shape argument strength" Arceneaux (2012, 272).

Why is it important to determine whether some types of argument are more successful than others? Classical critiques suggest that political rhetoric is generally and inherently damaging to democracy because it prioritises emotion and passion over reason, and inhibits rational deliberation between citizens (Elster, 1998). However, recent

work in normative political theory which attempts to "rehabilitate rhetoric" (Chambers, 2009; Dryzek, 2010) suggests that, while rhetoric may not be damaging *per se*, specific *forms* of rhetoric – particularly when used to communicate "vapid and vacuous" statements rather than substantive policy information – should still be viewed as a threat to deliberative ideals (Chambers, 2009, 337). If voters consistently respond to arguments that are low in informational content but rich in bombast and élan, we might worry that the quality of deliberation has fallen. By contrast, if voters are more consistently persuaded by arguments that reference relevant factual information and expert authority, we might have less concern.

In this paper, we provide the first quantitative evaluation of the relative effectiveness of a large number of different rhetorical elements across a large number of political issues by introducing a new experimental design and associated modelling approach. We examine types of arguments frequently made in contemporary British politics, and especially in speeches delivered in the UK parliament. The rhetorical elements that we identify relate to ongoing debates in diverse literatures in political communication, and are relevant to domestic politics in many countries. Our main experiment tests 336 individual arguments that use one of 14 distinct rhetorical elements to make arguments on each side of 12 policy issues in the UK. We present pairs of these arguments to survey respondents and ask them to assess which of the pair is most persuasive. We then use the distribution of responses to generate estimates of the relative persuasiveness of each of the arguments and, in turn, of the average persuasiveness of each of the rhetorical elements. A central virtue of our design is that, by presenting many implementations of each element, we are able to draw inferences about the relative effectiveness of different rhetorical strategies averaged across different political issues.

In addition to being a study of argument types and their relative persuasiveness, this paper is also a methodological argument for a different sort of experimental design.

Recent meta-analyses of persuasion field experiments (Kalla and Broockman, 2018) and online advertising experiments (Coppock, Hill and Vavreck, 2019) move beyond merely collecting existing study results towards fielding multiple similar experiments for the purpose of pooling evidence from them. Our design takes this logic much further. Researchers using survey experiments seldom want to test the effects of particular treatment texts on particular survey prompts. Rather, they typically want to make broader claims about a *latent treatment* (Grimmer and Fong, 2019) or treatment type, of which a treatment text is just one implementation. Many of the latent treatments that researchers wish to assess are likely to have variable effects across specific implementations. If we are interested in the *type* of treatment, rather than the specific treatment text, using many implementations rather than few or one should not wait for a meta-analysis of a mature research literature.

A traditional objection to this is that we would need to collect far larger samples to test many implementations of a latent treatment type. However, once we recognise that we are far less interested in the effects of specific treatment implementations than the *distribution* of such effects across implementations, we can use multilevel modelling to estimate this distribution using a large number of implementations, each of which would be statistically underpowered if analysed alone. In addition to reducing the risk that our conclusions about the latent treatment types will be confounded by the idiosyncrasies of single implementations, we illustrate how this approach enables post-experimental checks related to specific confounding concerns.

Our main substantive results reinforce the value of these methodological innovations. We find that there are modest average differences between different rhetorical element types. One of the strongest rhetorical elements in our experiment is *appeals to authority* – that is, arguments that seek support for an issue by reporting the view of an entity with relevant subject area expertise. The role of expertise in political debate be-

came a prominent issue in UK politics during the Brexit referendum in 2016 when a leading figure in the Leave campaign declared that the public "have had enough of experts".[1] Our results suggest that, despite this view, making appeals to authority remains among the most persuasive ways to argue about political issues. By contrast, the weakest arguments, on average, are those that employ *ad hominem* attacks and those that rely on *metaphor and imagery* to win support for a policy stance. While empirical evidence on the efficacy of negative attacks in political communication is mixed (Lau, Sigelman and Rovner, 2007), recent studies argue that the use of metaphor can be central to successful political campaigning (Charteris-Black, 2011) and a major determinant of the ways that individuals reason about politics (Thibodeau and Boroditsky, 2011). Our results build on both of these literatures, and suggest that when compared to many other common forms of political rhetoric, arguments of these types are relatively unpersuasive in the eyes of the UK public, *at least on average*.

However, and in some sense more importantly, we find that the heterogeneity in the effectiveness of specific implementations of these rhetorical elements is much larger than these average differences. While *appeals to authority* are more persuasive than other rhetorical styles on average, some appeals of this sort are still among the weakest arguments we test. Similarly, arguments that rely on making *comparisons to other countries* feature in the lists of the most and least persuasive in our experiment, depending on the specific implementation and issue. This finding represents an important lesson for the interpretation of existing studies of rhetorical effectiveness in political communication, a large number of which are based on experiments which relate to single policy issues. While it is not novel to observe that the effects of particular experimental implementations may not generalise to other domains, we directly quantify the substantial variance of the effects of the same treatment types across issues.

---

[1] Britain has had enough of experts, says Gove, Financial Times, 3 June, 2016

Ultimately, our goal is to understand which types of arguments induce voters to support or oppose policy proposals on different issues. However, *persuasion* of this sort is different from the self-reported judgements of argument *persuasiveness* that we elicit in our experiment (Vavreck et al., 2007; Graham and Coppock, 2019). To address this concern, we conduct a separate, out-of-sample validation experiment. We find that respondents' evaluations of which arguments are more *persuasive* in our initial experiment strongly predict the direction and magnitude of those arguments' ability to *persuade* different respondents to actually change their stated attitudes in the validation. The validation demonstrates large persuasion effects on average, but we again observe large variation in these treatment effects across policy issues. Therefore, in addition to providing an important check on the validity of our main experimental design and measurement strategy, the validation also reinforces our central methodological argument. Argument quality varies substantially and researchers should exercise caution when generalising the results of studies in specific policy areas to different issue domains.

Rhetoric, persuasion, and public opinion

Canonical work in the literature takes a broad view of what constitutes political rhetoric, seeing it as a "range of methods for persuading others" (Charteris-Black, 2011, 13). Politicians' arguments often share common rhetorical elements which are thought to be one source of their persuasive appeal, and we share the understanding of Atkins and Finlayson (2013, 161) that analyses of political rhetoric should focus "on the varied kinds of proof or justification found in political argument." Several existing typologies partition political arguments into a number of distinct rhetorical categories (eg Aristotle, c.322 BCE; Charteris-Black, 2011; Finlayson, 2007), but – as we describe below – our focus is on argument-types that arise regularly in UK politics.

Research into political rhetoric is not always described as such, but one goal of a large body of public opinion research is to measure the persuasive effects of different forms of political argument. The core conceptual focus of research in this field is whether (and to what degree) a given rhetorical element can persuade citizens to change their political views. For instance, though politicians may construct very different metaphors to argue about the economy (Barnes and Hicks, 2019), crime (Thibodeau and Boroditsky, 2011) and healthcare (Schlesinger and Lau, 2000), it might be the use of metaphor itself that is "essential to their persuasiveness." (Charteris-Black, 2011, 2). Existing work has considered the effects of a wide range of rhetorical elements on public opinion, including populist rhetoric (Atkins and Finlayson, 2013; Bos, Van Der Brug and De Vreese, 2013; Hameleers, Bos and de Vreese, 2017; Hameleers and Schmuck, 2017); negative or *ad hominem* attacks (Lau, Sigelman and Rovner, 2007); morality- and values-based appeals (Jung, Forthcoming; Nelson, 2004); appeals based on expected costs and benefits of policy (Jerit, 2009; Riker, 1990); and the use of expert cues and endorsements (Boudreau and MacKenzie, 2014; Dewan, Humphreys and Rubenson, 2014; Atkins and Finlayson, 2013).

Similarly, the literature on framing effects asks whether strategic language use by elites can change the factors that are relevant to voters' evaluations of policy options. Existing research in this area considers frames that emphasise free-speech concerns (Nelson, Clawson and Oxley, 1997), fiscal cost-benefit considerations (Leeper and Slothuus, 2018, 15), and the importance of civil liberties (Chong and Druckman, 2010), among others. Though these studies are often concerned with how political communications are portrayed in the mass media, they all engage with the idea that politicians can persuade voters to endorse particular policy options by using language strategically.

Our study addresses three limitations of the existing literature on rhetoric and persuasion. First, the cumulative evidence from these studies suggest that elite communication can substantially shift public opinion, and several authors express concern that

such results imply that citizens do not hold stable and well-formed preferences (Druck-man, 2004; Disch, 2011). Others have argued that there is an important role for rhetoric in the process of democratic deliberation (Dryzek, 2010), and that certain *forms* of rhetoric are more defensible than others (Chambers, 2009). Chambers (2009, 328), for example, emphasises that it is not rhetoric *per se* that is problematic, but specifically *plebiscitary* rhetoric—populist appeals divorced from factual merits—that represents a "threat to deliberation." By contrast, less problematic is *deliberative* rhetoric, which "makes people think, it makes people see things in new ways, it conveys information and knowledge, and it makes people more reflective" (Chambers, 2009, 335). A key goal for empirical studies, then, should be to determine whether different forms of rhetoric are differentially persuasive.

Unfortunately, the existing evidence on rhetoric and persuasion, which comes predominantly from survey experiments, provides little information regarding such comparisons. In almost all the papers cited above (and many others not cited) the persuasiveness of the relevant style or frame of interest – populism; metaphor; morality; etc – is evaluated in the context of vignette experiments where a treatment text containing the relevant element is contrasted with a control condition that does not include that element.[2] We are not the first to observe that *comparisons* of persuasiveness between the element of interest and other plausibly applicable rhetorical elements are very rare (Chong and Druckman, 2007a, 638; Sniderman and Theriault, 2004, 141). Thus, our first contribution is to provide novel evidence about which of a relatively large number of types of political rhetoric are more or less effective for shaping public opinion.

Second, the overwhelming majority of survey experiments which estimate the causal effects of persuasive speech do so in the context of a single-issue. Existing work on external validity in survey experiments has explored whether effects estimated from con-

---

[2]See, for example, Bos, Van Der Brug and De Vreese (2013), Arceneaux (2012), Jung (Forthcoming), Nelson (2004), Jerit (2009).

venience samples match those from representative samples (Berinsky, Huber and Lenz, 2012; Coppock, 2019), and whether experimental findings are replicated in comparable real-world settings (Barabas and Jerit, 2010; Bechtel et al., 2015). However, these external validity concerns are distinct from the idea that effects detected in an experiment on one issue may not generalise to a broader population of political issues for which politicians might use these types of rhetoric. As Druckman (2004, 685) suggests, "scholars need to carefully consider the context under study – perhaps, to an even greater extent than the population."

Some existing research (Lecheler, de Vreese and Slothuus, 2009; Hopkins and Mummolo, 2017) suggests that the persuasive effects of different frames vary across policy issues and it seems plausible *a priori* that certain types of rhetoric may be more appropriate for certain policies issues. For instance, are the legitimizing effects of populist rhetoric the same for issues relating to nuclear power (Bos, Van Der Brug and De Vreese, 2013) as they are for immigration? Are loss aversion arguments equally persuasive on economic issues as they appear to be on public health issues (Arceneaux, 2012)? Are rhetorical statements that make reference to "cost/benefit" considerations as influential when applied to issues of education as they are to issues of welfare (Jerit, 2009)? Of course, many studies in this literature make nuanced arguments about rhetorical effectiveness, paying close attention to the conditions under which different strategies are likely to be persuasive. Nevertheless, understanding whether some rhetorical elements are predictably more or less effective when considered across multiple policy issues remains an important and open question. Our second contribution, therefore, is to provide evidence of the distribution of effectiveness of rhetorical elements across a wide range of political issues.

Third, our approach also helps us to overcome a methodological problem that is common to vignette-style experiments which use single-text treatments. Grimmer and

Fong (2019) argue that latent treatments which are of interest to the researcher often co-occur with other textual features in experimental treatment texts. When this is true, effects estimated from such texts cannot necessarily be attributed to the latent treatment, as they might reflect instead the effects of these other correlated features.

In our design, which is similar to the design in Grimmer and Fong (2019), we provide *several* texts per latent treatment, thereby allowing background features which might confound our latent concepts of interest to vary. If background features vary independently of the concept of interest, then researchers can average over the effects of these separate treatments and attribute the average effect to the latent concept. Even if these potentially confounding background features do not vary independently of the concept of interest, having multiple treatment texts means that we are able to statistically control for any measurable confounding features of those texts. In combination, these aspects of our design mean that we can be much more confident that the treatment effects that we estimate in our experiment are attributable to the latent concepts (rhetorical elements) that motivate our study.

We build on the intuition and formalization presented in Grimmer and Fong (2019), but our study differs from theirs in terms of research design and empirical application. First, the central estimand for Grimmer and Fong is the average marginal component effect of their latent treatments, which they estimate using linear regression. By contrast, we illustrate how to use a multilevel modelling approach to characterise the *average* effect of each of our latent treatments as well as the *distribution* of the treatment effects across text implementations. Quantifying the variation in treatment effects across implementations provides important information about the generalisability of findings from existing single implementation studies. Second, while Grimmer and Fong provide evidence of the effectiveness of using multiple text-based implementations in the context of a single example, we apply this idea to a much larger set of latent treatments in

the field of political persuasion and document substantively important findings about the heterogeneity in argument strengths of different types.

Experimental design

We start by distinguishing between three concepts that are central to the structure of our experimental design: policy issues, rhetorical elements, and arguments. A **policy issue** refers to an issue that is subject to some level of political debate, where government could plausibly take action. In our setting, we focus on 12 policy issues in contemporary British politics: "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" and "Sugar tax in the UK" In deciding which policies to include, we focused on identifying those where there were clear political disagreements, both among politicians and the public, but where these divisions were not among the highest profile issues in British politics.

A **rhetorical element** is a feature of political argument that is used to emphasise the desirability or undesirability of a given policy. We based our categorisation of rhetorical elements on close reading of contemporary political debates. We began with a list of possible rhetorical categories, and then expanded and refined our categorisation by reading through transcripts of debates in the UK House of Commons and House of Lords that related to the issues defined above. Sourcing our arguments from parliamentary debates is helpful for situating our study in the context of real-world politics, and is consistent with calls to study "political arguments as they take place 'in the wild'" (Finlayson, 2007, 552). These debates provide a large repository of arguments about specific policy

areas, which tend to mirror those used by UK politicians in public speeches outside of parliament. The set of rhetorical elements that we evaluate, which was not intended to be exhaustive, is given in table 1.[3] While our primary goal is to quantify the persuasiveness of rhetorical appeals used in contemporary politics, our design is amenable to any arbitrary categorisation of arguments into types so long as the researcher is able to write multiple implementations of the same treatment concepts.

An **argument** is a text that makes a case in favour or against a specific policy. While real-world arguments sometimes include multiple rhetorical elements, for the purposes of our experiment we designed arguments that used a single element from the typology that we developed. A consequence of this decision is that our experiment is unable to evaluate whether certain rhetorical elements are more or less effective when used in conjunction with other elements. While interaction effects between elements are possible – and in some cases likely – we focus here on establishing the relative persuasiveness of our rhetorical elements when considered individually. For each policy issue, we wrote two separate arguments for each of the rhetorical elements: one arguing in favour of the policy, and one arguing against. This results in 14 x 12 x 2 = 336 separate arguments which are the basic treatments in our experiment.

To ensure the arguments we used resembled the types of argument used by politicians in the UK, we searched through the transcripts of UK parliamentary debates that pertained to the policy issues outlined above. From these debates, we extracted sentences and paragraphs that corresponded to our rhetorical elements, and then edited these texts into the form we use in the experiment. In the appendix we present all 336 arguments, and for many of the sentences we provide hyperlinks to the relevant source documents. When it was not possible to identify an example of our rhetorical styles in

---

[3]We considered further element types that feature in UK debate—such as "examples of personal narrative" and "appeals to freedom"—but found it too difficult to write treatments for them across all issues in the experiment.

Table 1: Elements

Element	Description
Appeal to fairness	A statement based on appeals to fairness. Uses the root "fair".
Costs vs benefits arguments	A statement which makes an explicit argument based on the costs and/or benefits of a policy. Uses the root "cost" and/or "benefit.
Country comparison	A statement is made about this policy or a similar policy in a named country, set of countries, or uses language about generic countries. This may be a statement of fact about whether the policy exists, or may be making an argument about its success/failure.
Crisis	A statement which emphasises the attractiveness or unattractiveness of a policy based on an argument that something is or is not a crisis. Must include the word "crisis".
Side-effects	A statement which emphasises the side-effects of a policy in order to persuade. Includes the phrase "unintended consequence/effect" or "side effect".
Metaphor/figure of speech	A statement which uses a figure of speech in which a word or phrase is applied to an object or action to which it is not literally applicable for rhetorical or vivid effect. May be hyperbolic.
Ad hominem	A statement which makes appeals based on undermining or impugning the motives of those on the other side of the argument. Might include mentions of corruption, ulterior motives, biased agendas, lack of consideration, hypocrisy, bad faith.
Appeals to expertise	A statement which reports the view of an entity with relevant subject area expertise in support of an argument. Explicitly mentions a not explicitly partisan entity – such as a professional body, academic organisation, research institute, think tank, union, business group, etc – by name.
Appeal to history	A statement of evidence from past policy experience *in the UK*. Includes explicit references to certain years and/or historical periods, or uses generic language about "the past" or "in previous years" or past generations.
Appeal to national greatness	A statement based on an appeal to national pride. Uses language about the UK being a world-leading country in this policy area, uses the word "great" as a descriptive, and/or makes explicit appeals to British values. Mentions the phrases "Britain" or "the UK" or "British" or "this country".
Appeal to populism	A statement which makes distinctions between elites and non-elites as the basis of a rhetorical appeal for the policy, situating the argument on the side of the non-elites. Does not require specific language to identify elites / non-elites, but can use familiar stand-ins to represent these categories.
Common sense	A statement which argues for or against a policy based on appeals to common sense or reasonableness. States that an argument for/against the policy is "common sense".
Morality	A statement which makes arguments for or against a policy based on morality concerns. Includes specific mentions of things being moral/immoral or right/wrong.
Public opinion	A statement which bases its argument on a claim about public opinion. Includes a phrase which has some quantifier (not necessarily numeric) about the support or opposition of the public for a policy.

the texts of the Commons' debates on a particular issue, we wrote arguments of our own, making the texts as similar as possible in style to those based on politicians' speeches.[4]

Survey instrument

We use these arguments as the basis of a forced-choice experiment which was fielded by YouGov to their UK online panel in June 2019. Following an introduction screen describing the task, respondents were presented with two arguments pertaining to a particular policy issue and asked which of the two arguments they found more persuasive.[5] Policy issues were sampled from the full set of 12 policy issues. For the selected policy, we then randomly sampled whether a respondent was presented with two arguments "in favour" of that policy (25% of responses), two arguments "against" that policy (25% of responses), or one argument "in favour" and one argument "against" (50% of responses).[6] We collected responses on four randomly selected issues from each of 3317 respondents, giving us a total of 13268 observations. An example prompt is given in Figure 1.

   As the wording of the survey prompt clearly reflects, this experiment assesses "persuasiveness" rather than "persuasion". That is, we look at *self-reported* assessments of arguments by respondents rather than the treatment effects of different arguments on respondents' own positions. Survey respondents tend to overestimate the effects of political stimuli on their own behaviour and attitudes (Vavreck et al., 2007; Graham and Coppock, 2019), and so we might be concerned that this approach will lead to over-

---

[4]In our analysis below, we show that we were no better or worse than UK politicians at writing persuasive political arguments.

[5]We use a paired-choice design for two reasons. First, by presenting respondents with competing arguments, our experiment more closely approximates the ways in which voters are exposed to political debate in the real world - a view that is shared by others (Chong and Druckman, 2007*b*, 102; Arceneaux, 2012, 271). Second, paired-profile designs of this sort are more successful, relative to single-profile designs, at generating estimates that replicate known real-world behavioural benchmarks (Hainmueller, Hangartner and Yamamoto, 2015).

[6]In the appendix, we show that "same side" and "different side" comparisons result in argument rankings that correlate at 0.81, which indicates that we can get nearly the same information from the different types of comparisons.

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 |

>

Figure 1: Experiment prompt

estimates of the variation in argument strengths. We address this issue by implementing an out-of-sample validation study (described below) where we check whether the arguments that respondents say are more persuasive are, in fact, better able to persuade people to endorse different policy positions.

Few prior studies provide comparisons of the effectiveness of different rhetorical elements, or comparisons of the effectiveness of the same element across multiple issues, and so we did not set out to test specific expectations about the relative persuasiveness of the elements we include in the experiment. Rather, we see our experiment as a measurement exercise that allows us to try to decompose the elements of a persuasive appeal that make it more or less effective. Our design, combined with the modelling framework outlined in the next section, allows us to estimate two key quantities relevant to this goal: the average (self-reported) persuasiveness of a wide variety of rhetorical elements, and the variation in persuasiveness of those elements across issues.

The Relative Persuasiveness of Rhetorical Elements

Modelling Persuasiveness

Our design generates responses that specify "winners" from a pairwise competition between arguments, with the possibility of ties.[7] Overall, we have J arguments, which we denote with $j = 1, ..., J$, and which we present to respondents, indexed as $i = 1, ..., N$, in paired comparisons. Our modelling task is to infer the efficacy of particular types of arguments given the results of the pairwise contests.

Our experiment results in an ordered response variable with three categories:

$$Y_i \in \begin{cases} 1 = \text{Argument 2 is more persuasive} \\ 2 = \text{About the same} \\ 3 = \text{Argument 1 is more persuasive} \end{cases} \tag{1}$$

To model this outcome, we adopt a variation on the Bradley-Terry model for paired comparisons (Bradley and Terry, 1952) where we model the log-odds that argument j beats argument j' in a pairwise comparison as:

$$log\left[\frac{P(Y_{j,j'} \leq k)}{P(Y_{j,j'} > k)}\right] = \theta_k + \alpha_j - \alpha'_j \tag{2}$$

where θ_k is the cutpoint for response category k and each argument j is described by a single "strength" parameter α_j. The strength parameter for a given argument, α_j, increases in the number of comparisons j "wins" against other arguments, and also in the strength of the arguments that j defeats. The intuition behind these parameters is straightforward: the stronger argument j is relative to argument j', the higher the prob-

---

[7]An alternative would be to ask respondents to *rate* different arguments on a common scale. Such an approach might result in more fine-grained information, but would come at the cost of decreased interpersonal comparability. The core of our approach would nonetheless apply to data from a rating task, with an appropriate multi-level model for an interval-level outcome.

ability that argument j beats argument j' in pairwise comparison.

If we had only a few arguments to test and a large number of responses involving each one, then we could simply use this as our full model specification and interpret the α_j parameters.[8] However, our primary quantity of interest is not the strength of individual arguments, but rather the *distribution* of the strength of arguments α_j for each of the 14 rhetorical elements. That is, we are interested in modelling how the strength of arguments vary as a function of rhetorical features that appear in those arguments. We therefore specify a hierarchical model for the α_j parameters.

Where $e(j) \in 1, \ldots, 14$ is the rhetorical element present in argument j, and $p(j) \in 1, \ldots, 12$ is the policy issue that the argument is about, and $s(j) \in 1, 2$ is the side of the issue that j argues for, we model the argument effects at a second-level using a model of the following form:

$$\alpha_j = \delta_{p(j),s(j)} + \mu_{e(j)} + \nu_j$$

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

$$\nu_j \sim N(0, \sigma_{e(j)}) \tag{3}$$

We assume a baseline effectiveness of arguments on the "for" versus the "against" side of each issue via the $\delta_{p(j),s(j)}$ parameters. These parameters separate the relative self-reported persuasive power of arguments from the degree to which respondents tend to agree with the side of the issue on which that argument appears. Note that, given the way that the α parameters enter equation 2, these parameters cancel in the case where both arguments in the pairwise comparison are on the same side of an issue. The next set of parameters $\mu_{e(j)}$ capture the average effect of each of our rhetorical elements. The final set of parameters are the ν_j, which are argument-specific "residuals" that characterise

---

[8]Each of the 336 individual treatments appears in an average of 79 pairwise comparisons in our data (sd = 8.6).

the distribution of argument-level effects around the element-type average. We estimate separate variance parameters for each element ($\sigma_{e(j)}$).[9] We estimate the model using Stan (Carpenter et al., 2017).[10]

The hierarchical model described by equation 3 distinguishes our approach from work on Canadian referendum arguments by Loewen, Rubenson and Spirling (2012), who also ask survey respondents to compare pairs of political arguments and estimate the persuasive power of those arguments via a structured Bradley-Terry model. In essence, Loewen, Rubenson and Spirling (2012) implement a version of equation 2, while we use equation 2 as the basis of a hierarchical model that better captures our primary quantities of interest. While the two modelling approaches are similar at the argument-level, our hierarchical model allows us to describe the distribution of element strengths across issues.

As we illustrate in the supplemental information, it is possible to recover very similar estimates for individual argument strengths—and nearly identical estimates for average rhetorical element strengths—using standard regression methods. We adopt a multilevel framework because the variance across treatments is a core quantity of interest, but we acknowledge that other analytic options are available for this kind of experiment.

Results

We present the main results from our model in Figure 2.[11] The figure shows the estimated average strength for each of our 14 rhetorical elements (μ_e) as well as for each of the 336 individual arguments ($\mu_{e(j)} + \nu_j$) that we include in the experiment. Blue numbers

---

[9]To identify the relative scale, $\delta_{p(j),1} = 0$ for all "against" arguments, and $\delta_{p(j),2}$ are estimated with a uniform prior for all "pro" arguments. We use uniform priors on the ω and σ_e parameters as well.

[10]We used 4 chains of 1000 iterations, after a 200 iteration burn-in. We present convergence diagnostics in the online appendix.

[11]In the appendix we report the $\delta_{p(j),s(j)}$ parameters which show that that there is substantial variation in the degree to which respondents think that arguments are persuasive as a function of which side of which issue they are on.
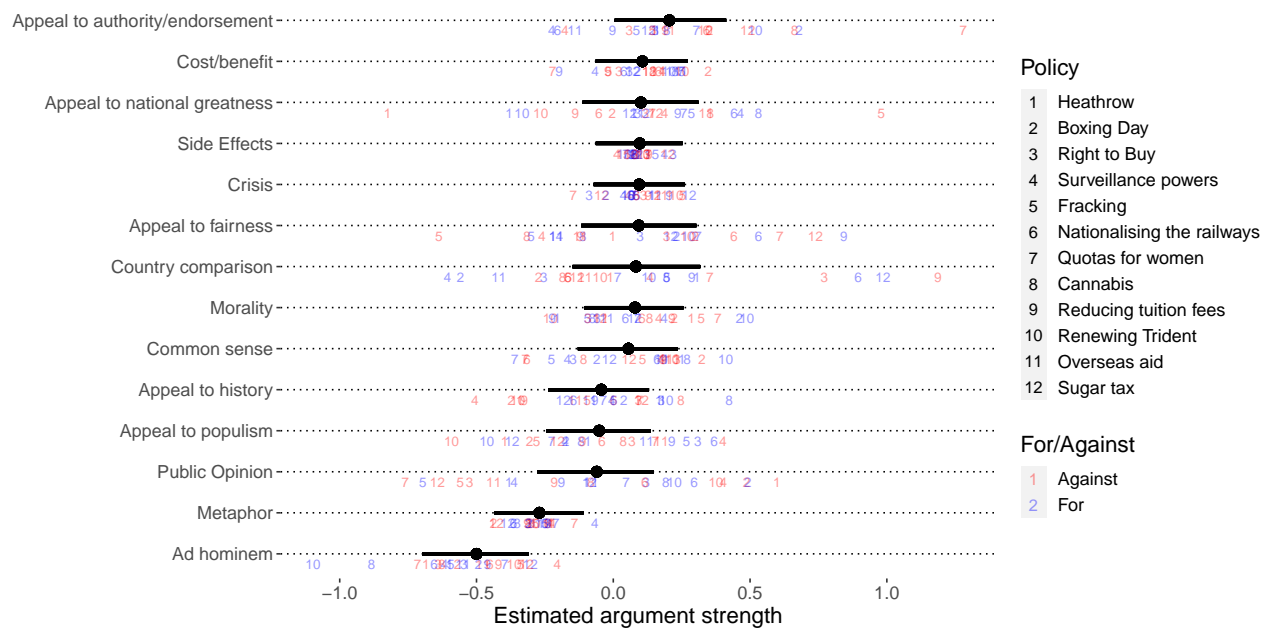
Figure 2: Estimated argument strengths

indicate arguments on the "for" side of the relevant issue, and red numbers indicate arguments on the "against" side of the issue. The numbers themselves relate to the different policy areas, which are listed in the legend.

Two main patterns of interest arise in Figure 2. First, there is some variation in the average self-reported persuasive power of our 14 rhetorical elements. The estimates suggest that respondents have a clear aversion to arguments based on ad hominem attacks impugning the characters or motives of those on the opposite side of the issue, as well as to arguments that are based on metaphor and imagery. Previous research has argued that metaphors have large effects on how individuals reason about solutions to social problems like crime (Thibodeau and Boroditsky, 2011) and also appear to help individuals in developing understanding of politics and public policy more generally (Schlesinger and Lau, 2000; Bougher, 2012). Some authors see metaphor as so central to the process of modern political communication that, for many politicians, "metaphor is essential to their persuasiveness" (Charteris-Black, 2011, 2). Our findings, by contrast, suggest that metaphor-based arguments are less persuasive on average than most other

types of rhetorical appeal that we evaluate.

The differences between the other element types are more modest, and it is difficult to be confident about their relative average strength.[12] The posterior probability that the mean strength of arguments based on appeals to authority and expertise is the highest among all element types is 0.53, versus the uniform prior probability of 0.07. We can be reasonably confident that some of the element types are stronger, on average, than others. For example, the posterior probability that appeals to authority are on average more effective is at least 0.9 versus arguments employing appeals to common sense, historical comparisons, populist arguments, appeals to public opinion, metaphors and ad hominem attacks. Similarly, the probability that populist appeals are, on average, less persuasive than to appeals to authority, costs and benefits, side effects, fairness, national greatness, and crisis is approximately 0.9 in each case.[13] Taken together, while the average differences between elements are modest, voters appear to find statements that include references to expertise ("Appeal to authority") and factual argument ("Cost/benefit", "Side effects") more convincing than statements that employ striking language but are thinner in terms of substantive policy-relevant content ("Ad hominem", "Metaphor", "Appeal to populism"). Given that normative concerns about rhetoric center on types of argument that are dedicated "first and foremost to gaining support for a proposition and only secondarily with the merits of the arguments" (Chambers, 2009), the ranking we uncover provides a relatively optimistic view of the rhetoric that is deemed persuasive by the UK public.

---

[12]The modest average differences between elements do not appear to be due to lack of engagement. First, we find considerable heterogeneity in persuasive power at the *argument* level (see below), indicating that respondents do distinguish between more and less persuasive arguments. Second, although in 28% of comparisons respondents indicated that the two arguments presented were "about the same", very few respondents gave this intermediate response across multiple comparisons. Only 6% of respondents gave 3 (out of 5) "about the same" responses, 5% gave 4 such responses, and we never observe a respondent giving 5 such responses. This suggests that the vast majority of respondents were sufficiently attentive to the task that they were able to adjudicate on the persuasiveness of different arguments in most cases.

[13]We report all possible pairwise comparisons in the appendix.

Second, figure 2 also clearly illustrates that there is a substantial degree of heterogeneity in the performance of arguments using the same rhetorical element. This is particularly so for certain element types. For example, statements using country comparisons to argue in favour of nationalising the railways and implementing a sugar tax, and also those arguing against extending the right to buy and reducing tuition fees, are among the most persuasive in our data. By contrast, other arguments of the same type – country comparisons arguing in favour of extending surveillance powers and closing large stores on Boxing Day – are amongst the weakest that we include in the experiment. Similarly, while appealing to national greatness to oppose fracking is a relatively persuasive way to argue, opposing the expansion of Heathrow using similar appeals is not. Further, argument strength heterogeneity is not equal across all element-types. For instance, the "metaphor" arguments tend to perform similarly to one another and the same is true of "crisis" and "side effect" arguments.

It is important to recognise that these are statements about the treatments that we tested, which may or may not reflect broader populations of arguments that one might define. It might be that we, or the MPs whose statements we adapted, are bad at ad hominem attacks, but that such attacks are effective when deployed more competently. Alternatively, it may be that certain forms of rhetoric – such as the use of metaphor – are less effective in written form than they would be if spoken aloud. Nonetheless, our finding of very substantial heterogeneity in the performance of different arguments using the same element type is unlikely to be very sensitive to these concerns. Moreover, all of these criticisms also apply to existing experiments that use single-text implementations of political communication styles. In some contexts, researchers are clear that their interest is in the efficacy of certain rhetorical elements as they pertain to specific policy areas,[14] but authors frequently aim to make more general claims about the

---

[14]See, for example, Barnes and Hicks (2019) and Feldman and Hart (2016).

persuasiveness of a given rhetorical element on the basis of experiments that provide evidence from only one or a few policy domains. The conclusion we draw from this analysis is that experimental estimates of the effects of rhetorical styles are likely to vary considerably in both sign and magnitude depending on the specific policies to which they relate.

Controlling for Argument-Level Confounders

Implementing multiple texts per latent treatment of interest helps to account for confounding by other text features by allowing us to average over the varying effects of those other features. However, this will only recover unbiased estimates of the latent treatment effects of interest if the variation in text features is uncorrelated with the latent treatments. Our design allows us to further mitigate this problem in cases where we can directly measure the background features that are a cause for concern. Because we have hundreds of treatment implementations, and a model for the effectiveness of these individual treatments, we can control for potential confounding features when estimating the element effects. We adapt equation 3 to include a vector of K argument-level measures, which we denote $x_{k,j}$:

$$\alpha_j = \delta_{p(j),s(j)} + \mu_{e(j)} + \sum_{k=1}^{K} \gamma_k x_{k,j} + \nu_j$$
$$\mu_e \sim N(0, \omega)$$
$$\nu_j \sim N(0, \sigma_{e(j)}) \tag{4}$$

The parameters γ_k represent conditional average linear effects of text-feature k on argument strength. We have identified seven argument-level variables which represent features of our argument texts that might plausibly confound the effects of the rhetor-
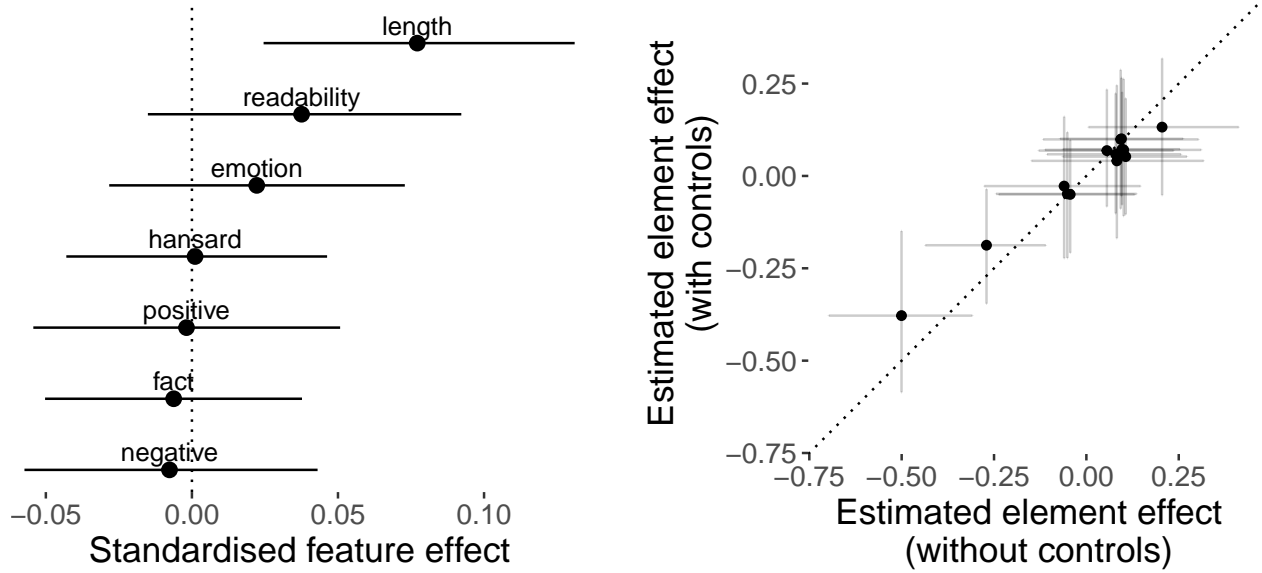
Figure 3: Control variable coefficient estimates (left) and comparison of element average effects with and without controls (right)

ical elements that we aimed to study: argument length, readability, positive and negative tone, overall emotional language, fact-based language, and whether the argument was based on parliamentary speech from Hansard or was created by the authors of this study.[15]

Figure 3 presents the results. The left-hand panel of the plot shows the standardised posterior point estimates and intervals of the γ_k parameters from equation 4, and the right-hand panel compares the point estimates of the element average parameters $\mu_{e(j)}$ from the models with (equation 3) versus without controls (equation 4).[16]

Of the 7 argument-level control variables we include, only length has a clearly significant effect on argument persuasiveness. On average, respondents find arguments with more words somewhat more persuasive than arguments with fewer words. We find

---

[15]Length is measured as the number of words in the argument, and readability using the Flesch's Reading Ease Score (Flesch, 1948). We measure tone using the proportion of words in each argument listed in the positive and negative categories of the Affective Norms for English Words dictionary (Nielsen, 2011); emotion using the 'affect' category from the 2015 Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker, Francis and Booth, 2001)); and fact-based language using in the 'quantitative' and 'numeric' categories of the LIWC dictionary.

[16]The dotted line in the right-hand panel of the figure is the 45 degree line.

weak evidence that readability and emotional content positively predict persuasiveness. There is no difference in average quality between the arguments written by the authors of this study and those from the parliamentary record.

The right-hand panel of the figure demonstrates that controlling for the additional text features has limited consequences for the estimated rhetorical style effects. There is a slight attenuation of the differences between the rhetorical styles because the two least popular element types had arguments that were somewhat shorter than average. Thus we find little evidence of what is conventionally called confounding (which Grimmer and Fong (2019) call "aliasing") in this context. The differences in the performance of our text treatments based on different rhetorical elements cannot be explained away by these measurable differences in the implementation of those elements in our text treatments. A major strength of our design is that we are able to assess robustness to alternative explanations for differences in the performance of different textual treatments after the experiment is completed.

In the appendix, we illustrate alternative multilevel models that we can fit to these data. We show that the strength of arguments is generally positively correlated across respondents of different age, education, attention to politics, and even past vote: arguments tend to be more/less effective for everyone, with limited heterogeneity across groups. With another variation on this model we show that the relative strength of arguments is similar when compared to arguments on the same side of the issue as they are when compared to arguments on the other side of the issue. With a third variation we show that the pro and con arguments on the same issue-element have correlated efficacy: some argument types may be a better match for some issues, regardless of which side the argument is made for.

From Persuasiveness to Persuasion

Our estimates of argument strength are based on responses to questions which prompt individuals to assess which arguments they find to be more persuasive. Do responses to these self-assessment questions (and the modelling approach that we apply to them) in fact identify arguments that, when delivered out of sample, actually persuade new respondents to endorse different policy positions? To answer this question, we fielded a validation experiment with YouGov to new respondents in February 2020 – eight months and one general election after the initial experiment – which we use to evaluate whether comparisons of argument persuasiveness translate into arguments that are more effective at changing opinion.

Design

In this validation experiment, our treatments are constructed from the arguments that we used in our initial experiment. Rather than presenting single arguments on each side, we present paragraphs constructed from multiple arguments. This sets a more demanding standard for validation: the differences in argument strength have to be robust to being presented with other arguments, and not to merely apply when the argument is presented alone.

 Using our estimates of argument strength ($\mu_{e(j)} + \nu_j$) we select the three most and three least persuasive arguments in favour and against each policy, which we concatenate to form short treatment paragraphs. As shown in Figure 4, each respondent sees two opposing paragraphs: one with arguments in favour of the policy, and one with arguments against the policy. For each policy area, we define two treatment conditions. In the "strong in favour" condition, respondents see the strongest arguments in favour of a policy and the weakest arguments against (according to the estimates from the initial

experiment). In the "strong against" condition, respondents see the weakest arguments in favour of a policy, and the strongest arguments against. We collected responses on two randomly selected issues from each of our 6600 respondents, giving us a total of 13200 responses.

As we show in Figure 3, argument strength is correlated with sentence length. If our treatment paragraphs here always used equal numbers of arguments, then the strong paragraphs would contain more words on average than the weak paragraphs and we might be concerned that differences in length might confound the effects of our latent quantity of interest, argument strength. Once again applying the idea that having some measurable variation in treatment texts allows us to address potential confounding, we randomly vary whether each paragraph is made up of all three, or just two of the three strongest/weakest arguments from our initial experiment. Our analysis presented below averages over this variation. A further multilevel model analysis presented in the appendix confirms that the conclusions hold when we control for the number of arguments and number of words in each paragraph.

A key difference between this validation experiment and our original experiment is that here we do not ask respondents to identify persuasive arguments, but instead directly ask "Are you for or against <policy issue>?", and respondents can select "For", "Not sure" or "Against". If we have truly identified persuasive sets of arguments, we should see a greater fraction of respondents endorsing the policy among those who see strong arguments in favour of that policy, and a smaller fraction endorsing the policy among those who see strong arguments against the policy.

In addition, the estimates from our initial experiment suggest bigger differences in the persuasiveness of argument sets in some policy areas than in others. In some policy areas we observe very strong strong arguments, and very weak weak arguments, while in other policy areas we observe only moderately strong strong arguments and moderately

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.

For	Against
It is just common sense that an airport as congested as Heathrow should be expanded. Expansion at Heathrow will bring real benefits across the country, including a boost of up to £74 billion to passengers and the wider economy, and these will easily surpass the costs of expansion.	Great nations don't waste money on vanity projects, and the expansion of Heathrow would be nothing more than a project of national vanity. Expanding Heathrow will enrich a private foreign-owned business at the expense of higher fares for ordinary passengers.

Are you for or against building a third Runway at Heathrow?

For	Not sure	Against

>

Figure 4: Experiment 2 prompt

weak weak arguments. In this validation experiment, when we pair strong arguments and weak arguments in each policy area, we have *ex ante* variation in expected treatment effect sizes: the magnitude of the treatment effects in experiment two should correlate positively with the expected difference in treatment strengths across policy areas as measured from experiment one.

To evaluate this expectation, we define the expected difference in argument strength for sets of three arguments on either side of a policy issue as the average of the strengths of the individual arguments:

$$\pi = \left(\frac{1}{3} \sum_{j \in \text{in favour}} \mu_{e(j)} + \nu_j \right) - \left(\frac{1}{3} \sum_{j \in \text{against}} \mu_{e(j)} + \nu_j \right) \tag{5}$$

For each policy issue, this calculation yields one relative strength value for the treatment condition where the "in favour" arguments are the strongest in our data and the

"against" arguments are the weakest ($\pi_{\text{strong in favour}}$), and one where the "against" arguments are strong and the "in favour" arguments are weak ($\pi_{\text{strong against}}$). The expected strength of the treatment when we compare the "strong in favour" and the "strong against" conditions is therefore given by the difference between these two quantities: $\pi_{\text{strong in favour}} - \pi_{\text{strong against}}$. This quantity will be larger for those policy issues where our initial analysis revealed greater variation in treatment strength across both in favour and against arguments.

Results

The results are given in figure 5. There are three main conclusions from this analysis. First, the left-hand panel of figure 5 shows the simple difference in mean support for each policy between respondents in the "strong in favour" condition and the "strong against" condition. The point estimates of the treatment effects are positive for every policy area, and most of them are significantly different from zero.

Second, many of the treatment effects we estimate are very large. The largest treatment effect we estimate is for the "Sugar tax" issue, where respondents in our "strong in favour" condition are 19 (95% interval: 13-24) percentage points more likely to endorse the policy than respondents in our "strong against" condition. Similarly, for the "Boxing day", "Quotas for women", "Cannabis", "Nationalising the railways", and "Fracking" issues, our point estimates imply that the strong arguments we deploy in favour of those policies persuade more than ten percentage points of respondents to endorse the policy relative to when we deploy strong arguments against the policy. The large size of these effects suggest that the experimental design and modelling strategy that we describe is successful in measuring the relative persuasive power of different political arguments.

Third, the treatment effects vary considerably in magnitude across policy issues. The right panel of figure 5 provides very strong evidence that expected differences in ar-
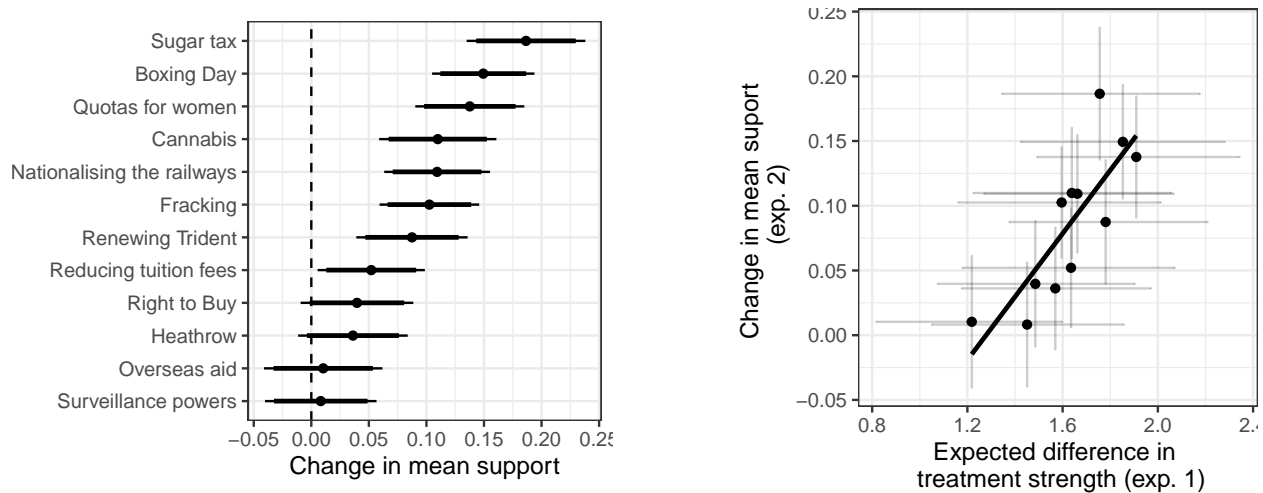
Figure 5: Experiment two results

gument strength from experiment one predict the magnitude of the treatment effects from experiment two. The y-axis measures the change in mean support for each policy area from our experiment, and the x-axis measures the expected difference in treatment strength based on our estimates from our first experiment (equation 5). These quantities are clearly positively related across the 12 issue areas, and the linear association between the two is significantly different from zero ($t = 4.4$). Larger interval differences on our argument strength scale measured in experiment one translate into larger persuasion effects when tested out of sample in experiment two.

Our measure of the expected difference in treatment strengths implicitly assumes that the effects of arguments as measured in our initial experiment are *additive*. However, respondents may not evaluate combinations of arguments on the basis of their average persuasive appeal. For instance, respondents may be persuaded by the set of arguments that contains the single strongest argument. Our experiment is not designed to test whether alternative argument aggregation rules of this sort would be more predictive of persuasion but this is an important avenue for future work.

Conclusion

Identifying which general argument forms are most persuasive is a long-standing goal in the study of politics. For Aristotle ([c.322 BCE](), 1355b), rhetoric entails "discovering the possible means of persuasion in reference to any subject whatsoever." However, though contemporary empirical work has established the persuasive effects of certain argument types in certain circumstances, scholarship on persuasion has not clearly decomposed the sources of persuasive appeal into distinct rhetorical elements. The central substantive finding of our study suggests why this might be the case: it is very difficult to identify rhetorical strategies that are *consistently* more persuasive than others when considered across multiple policy issues. Basing our design on the types of rhetoric that are regularly found in real-world political speeches in the UK, we found only moderate average differences in the persuasive power of 14 different rhetorical elements, and we demonstrated that there is significant heterogeneity in argument strength within element types. Together, these findings imply that the persuasiveness of different argument types is likely to be highly context-dependent, and that analysing the rhetorical structure that characterises an argument allows us to predict the persuasiveness of that argument only to a limited extent.

Our empirical findings also reinforce two key methodological points. First, they imply external validity concerns for existing studies that rely on single implementations of latent treatments in texts. Though some existing persuasion studies make issue-specific claims, many seek to offer more general conclusions about the effectiveness of different forms of rhetoric. The effect heterogeneity we uncover for arguments of the same rhetorical type suggests that the persuasive power of a given rhetorical element may be very different across issues and therefore that such generalisations should be made with caution. Second, they suggest that researchers should more generally use designs based

on pooling evidence from many small implementations rather than few large ones.

In making this point about the external validity of other studies, we are arguably guilty of the same kind of extrapolation that we are cautioning against. In a narrow sense, we have demonstrated that arguments of the types frequently used in the UK Parliament vary widely in their ability to persuade UK citizens, across a set of medium salience UK political issues. Does this translate to other kinds of survey experiments that political scientists use to assess theories of public opinion and political psychology? We cannot clearly demonstrate that it does. Nonetheless, we think that our empirical results usefully demonstrate a general theoretical concern, which clearly applies *as a concern* across a wide range of studies. Some experimental domains may not exhibit this level of implementation-level heterogeneity, for various reasons. But, at the very least, a strong theoretical argument ought to be expected when researchers move from their specific experiment to more general claims about an underlying phenomenon. Better than such an argument though, would be more widespread use of the core approach of this paper: conducting a larger number of smaller experiments, and using multilevel models to characterise the distribution of results.

Our design and results suggest several new avenues of research into the persuasive effects of different rhetorical strategies. First, our experiment employs written texts but rhetorical skill may manifest differently in spoken and written forms. It is plausible that the ordering of elements that we describe would change if we used videos of politicians speaking rather than texts of their speeches as the basis of our experiment. Using multiple treatment implementations to capture latent treatment effects would apply equally well to video-based as to text-based treatments. Additionally, video treatments would allow researchers to assess a wider variety of rhetorical elements which are difficult to capture adequately in written form.

Second, our analysis has implicitly assumed that the persuasive effects of rhetorical

elements are additive, but this may miss important interactions between elements. For instance, observers of contemporary political debates might believe that *ad hominem* attacks are likely to be more effective when combined with populist appeals of different types. Our design gives a framework for evaluating such hypotheses. Investigating whether and how interactions of this sort affect respondent choices would be informative about how people process rhetorical appeals, and could help to reveal how arguments could be combined to optimize their persuasive effects.

Third, interactions may also exist between elements and issues. While we have focused on establishing the average persuasiveness of rhetorical elements *across* issues, researchers could use our experimental framework to begin the process of accumulating evidence on the effectiveness of different argument types *within* specific policy domains.

Fourth, our estimates reflect the effects of only short-run exposure to different types of rhetoric. An interesting further development of the findings we present here would be to embed our experimental design in a panel study, which would allow researchers to evaluate how persuasion effects vary as voters are exposed to rhetorical strategies over a longer period of time.

Finally, our model characterises the strength of a set of rhetorical elements that we defined *ex-ante*. Future work might focus on *discovering* sets of textual features that are predictive of persuasiveness, a task that might be better achieved through a different modelling approach. For instance, when researchers have a corpus of pre-existing texts, they might characterise those texts with a very large number of features, such as the presence and absence of particular words, or measures of textual complexity, or a distribution over topics. In such a case, working out which of the (potentially very many) features are most predictive of persuasion would be the key challenge, and so a more flexible, regularized model – such as ridge or lasso regression – would be an appropriate choice.

References

Arceneaux, Kevin. 2012. "Cognitive biases and the strength of political arguments." *American Journal of Political Science* 56(2):271–285.

Aristotle, ([undated]. c.322 BCE. *Rhetoric.* Kessinger Publishing, 2004.

Atkins, Judi and Alan Finlayson. 2013. "'… A 40-year-old black man made the point to me': Everyday knowledge and the performance of leadership in contemporary British politics." *Political Studies* 61(1):161–177.

Barabas, Jason and Jennifer Jerit. 2010. "Are survey experiments externally valid?" *American Political Science Review* 104(2):226–242.

Barnes, Lucy and Timothy Hicks. 2019. "The Household Finance Analogy Does Not Drive Mass Support for Austerity." *Working Paper* .

Bechtel, Michael M, Jens Hainmueller, Dominik Hangartner and Marc Helbling. 2015. "Reality bites: The limits of framing effects for salient and contested policy issues." *Political Science Research and Methods* 3(3):683–695.

Berinsky, Adam J, Gregory A Huber and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political analysis* 20(3):351–368.

Bos, Linda, Wouter Van Der Brug and Claes H De Vreese. 2013. "An experimental test of the impact of style and rhetoric on the perception of right-wing populist and mainstream party leaders." *Acta Politica* 48(2):192–208.

Boudreau, Cheryl and Scott A MacKenzie. 2014. "Informing the electorate? How party cues and policy information affect public opinion about initiatives." *American Journal of Political Science* 58(1):48–62.

Bougher, Lori D. 2012. "The case for metaphor in political reasoning and cognition." *Political Psychology* 33(1):145–163.

Bradley, Ralph Allan and Milton E Terry. 1952. "Rank analysis of incomplete block designs: I. The method of paired comparisons." *Biometrika* 39(3/4):324–345.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2017. "Stan: A probabilistic programming language." *Journal of statistical software* 76(1).

Chambers, Simone. 2009. "Rhetoric and the public sphere: Has deliberative democracy abandoned mass democracy?" *Political theory* 37(3):323–350.

Charteris-Black, Jonathan. 2011. *Politicians and rhetoric: The persuasive power of metaphor.* Springer.

Chong, Dennis and James N Druckman. 2007*a*. "Framing public opinion in competitive democracies." *American Political Science Review* 101(4):637–655.

Chong, Dennis and James N Druckman. 2007*b*. "A theory of framing and opinion formation in competitive elite environments." *Journal of communication* 57(1):99–118.

Chong, Dennis and James N Druckman. 2010. "Dynamic public opinion: Communication effects over time." *American Political Science Review* 104(4):663–680.

Coppock, Alexander. 2019. "Generalizing from survey experiments conducted on Mechanical Turk: A replication approach." *Political Science Research and Methods* 7(3):613–628.

Coppock, Alexander, Seth J Hill and Lynn Vavreck. 2019. "Persuasive Effects of Presidential Campaign Advertising: Results of 53 Real-time Experiments in 2016.".

Dewan, Torun, Macartan Humphreys and Daniel Rubenson. 2014. "The elements of political persuasion: Content, charisma and cue." *The Economic Journal* 124(574):F257–F292.

Disch, Lisa. 2011. "Toward a mobilization conception of democratic representation." *American political science review* 105(1):100–114.

Druckman, James N. 2004. "Political preference formation: Competition, deliberation, and the (ir) relevance of framing effects." *American Political Science Review* 98(4):671–686.

Dryzek, John S. 2010. "Rhetoric in democracy: A systemic appreciation." *Political theory* 38(3):319–339.

Elster, Jon. 1998. "Deliberation and constitution making." *Deliberative democracy* 97:111.

Feldman, Lauren and P Sol Hart. 2016. "Using political efficacy messages to increase climate activism: The mediating role of emotions." *Science Communication* 38(1):99–127.

Finlayson, Alan. 2007. "From beliefs to arguments: Interpretive methodology and rhetorical political analysis." *The British Journal of Politics and International Relations* 9(4):545–563.

Flesch, Rudolph. 1948. "A new readability yardstick." *Journal of applied psychology* 32(3):221.

Graham, Matthew and Alexander Coppock. 2019. "Asking About Attitude Change." *Working Paper* .

Grimmer, Justin and Christian Fong. 2019. "Causal Inference with Latent Treatments." *Working Paper* .

Hainmueller, Jens, Dominik Hangartner and Teppei Yamamoto. 2015. "Validating vignette and conjoint survey experiments against real-world behavior." *Proceedings of the National Academy of Sciences* 112(8):2395–2400.

Hameleers, Michael and Desirée Schmuck. 2017. "It's us against them: A comparative experiment on the effects of populist messages communicated via social media." *Information, Communication & Society* 20(9):1425–1444.

Hameleers, Michael, Linda Bos and Claes H de Vreese. 2017. ""They did it": The effects of emotionalized blame attribution in populist communication." *Communication Research* 44(6):870–900.

Hopkins, Daniel J and Jonathan Mummolo. 2017. "Assessing the breadth of framing effects." *Quarterly Journal of Political Science* 12(1):37–57.

Jerit, Jennifer. 2009. "How predictive appeals affect policy opinions." *American Journal of Political Science* 53(2):411–426.

Jung, Jae-Hee. Forthcoming. "The Mobilizing Effect of Parties' Moral Rhetoric." *American Journal of Political Science* .

Kalla, Joshua L and David E Broockman. 2018. "The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments." *American Political Science Review* 112(1):148–166.

Lau, Richard R, Lee Sigelman and Ivy Brown Rovner. 2007. "The effects of negative political campaigns: a meta-analytic reassessment." *Journal of Politics* 69(4):1176–1209.

Lecheler, Sophie, Claes de Vreese and Rune Slothuus. 2009. "Issue importance as a moderator of framing effects." *Communication research* 36(3):400–425.

Leeper, Thomas J and Rune Slothuus. 2018. "Can citizens be framed? How information, not emphasis, changes opinions." *Working paper* .

Loewen, Peter John, Daniel Rubenson and Arthur Spirling. 2012. "Testing the power of arguments in referendums: A Bradley–Terry approach." *Electoral Studies* 31(1):212–221.

Nelson, Thomas E. 2004. "Policy goals, public rhetoric, and political attitudes." *The Journal of Politics* 66(2):581–605.

Nelson, Thomas E, Rosalee A Clawson and Zoe M Oxley. 1997. "Media framing of a civil liberties conflict and its effect on tolerance." *American Political Science Review* 91(3):567–583.

Nielsen, Finn Årup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." *arXiv preprint arXiv:1103.2903* .

Pennebaker, James W, Martha E Francis and Roger J Booth. 2001. "Linguistic inquiry and word count: LIWC 2001." *Mahway: Lawrence Erlbaum Associates* 71(2001):2001.

Rhetorica ad herennium. c.80 BCE. *Rhetorica ad herennium*.

Riker, William H. 1990. "Heresthetic and rhetoric in the spatial model." *Advances in the spatial theory of voting* 46:50.

Schlesinger, Mark and Richard R Lau. 2000. "The meaning and measure of policy metaphors." *American Political Science Review* 94(3):611–626.

Sniderman, Paul M and Sean M Theriault. 2004. "The structure of political argument and the logic of issue framing." *Studies in public opinion: Attitudes, nonattitudes, measurement error, and change* pp. 133–65.

Thibodeau, Paul H and Lera Boroditsky. 2011. "Metaphors we think with: The role of metaphor in reasoning." *PloS one* 6(2):e16782.

Vavreck, Lynn et al. 2007. "The exaggerated effects of advertising on turnout: The dangers of self-reports." *Quarterly Journal of Political Science* 2(4):325–343.

Supplemental Information - The Variable Persuasiveness of Political Rhetoric

Jack Blumenau   *University College London*
Benjamin E Lauderdale   *University College London*

Contents

Response Data by Element and Policy

The plots in this section show the raw response distribution for the main experiment. Each small point is a single argument on one side of a single issue, the large points show averages by element (top panel) and by policy, for and against (bottom panel)

Argument success by element
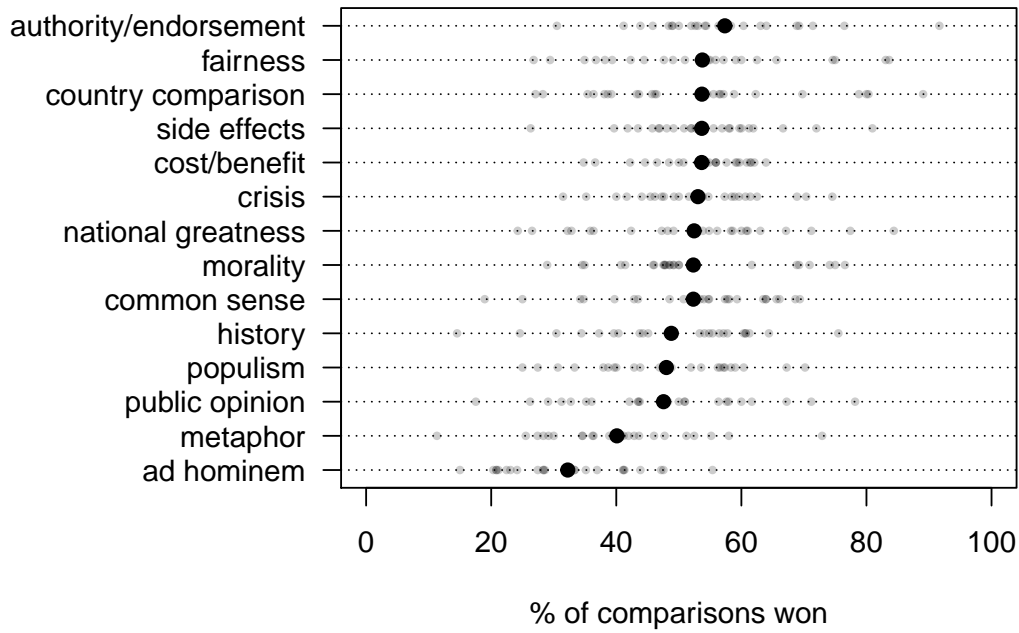
Argument success by policy

Figure A1: Raw data

Sensitivity to modelling assumptions

In this section, we investigate the sensitivity of our key estimates to changes in modelling assumptions.

Argument-level estimates

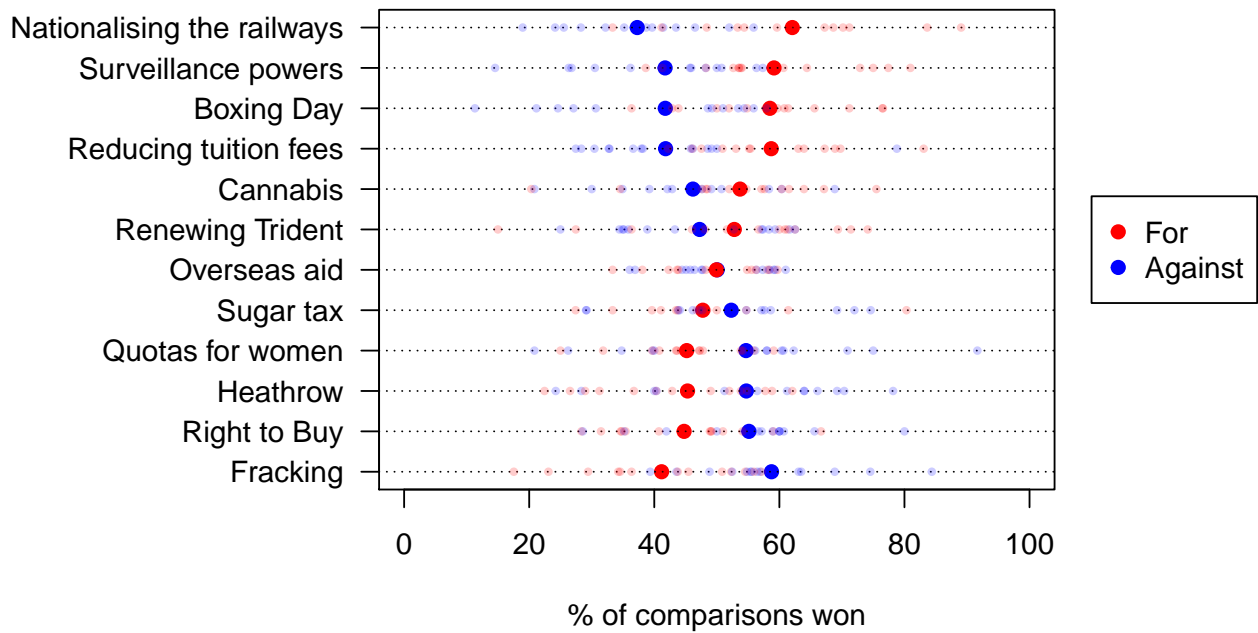First, we estimate the persuasiveness of each of our 336 arguments without relying on the multilevel structure described in equations 2 and 3 of the paper. Instead, we stack our data such that we have two observations for each pairwise comparison in our data - one observation for each of the arguments involved in the comparison. Our outcome, $Y_{i(j,r)}$, is 1 (-1) if respondent r rates argument j as more (less) persuasive in a given comparison, and 0 if the respondent rates both arguments in the comparison as "About the same". We then model this outcome with linear fixed-effect regressions of the following form:

$$Y_{i(j,r)} = \alpha_j + \epsilon_{i(j,r)} \tag{A1}$$

where the α_j parameters are fixed-effects for each of the 336 arguments. We estimate this model using OLS, omitting the intercept, and clustering standard errors at the respondent level. Figures A3 and A4 present these estimates both by element and by policy area.

We note that equation A1 differs in several ways from the the multi-level modelling strategy we employ in the paper. First, here we are modelling responses on a linear outcome scale, where in the paper we model the outcome according to an ordered-logistic distribution. As a consequence, the magnitudes of the coefficient estimates from this analysis and those from the multilevel model are not directly comparable. Second, the Bradley-Terry model we use in the paper controls for the strength of opposing arguments, while equation A1 relies solely on the randomization of arguments to respondents to generate unbiased estimates of argument strength. Finally, here argument strengths are estimated via fixed-effects, while the estimates we present in the paper are subject to a degree of regularisation due to the assumptions of the multilevel model. Despite these modelling differences, the correlation between these estimates

to those obtained from our multilevel model is 0.84. We plot the fixed-effect and multilevel model estimates against each other in the left-hand panel of figure A5.

One potential concern with our results is that, because we only observe each argument a limited number of times, we may not be adequately powered to detect systematic variation in argument strength. However, these figures reveal that, despite the fact that each individual argument only appears in relatively few pairwise comparisons (79 on average), there are nevertheless large and significant differences in the relative persuasiveness of many of these arguments. The F-test comparing the model in equation A1 to a null model where arguments all perform equally well in the broader population allows us to reject the hypothesis that the estimated differences in argument strengths that we report are attributable to sampling variation alone ($F = 5.91$, $df_1 = 1$ $df_2 = 336$, $p = 0.016$).

We can also compare the distribution of fixed-effects that we recover from our experiment to the distribution that would result from a set of randomly-imputed responses. To do so, we provide results from a simulation exercise in which we randomly permute the responses from our experiment and then estimate equation A1. Repeating this process 500 times, we then calculate the standard deviation of the α_j parameters, $\hat{\sigma}_{\alpha_j}$, and plot the distribution of this quantity in figure A2. The vertical red line indicates the standard deviation of the α_j parameters as calculated from the real experimental data. If we were unable to recover any predictable variation in argument strength, the standard deviation for the fixed-effects that we estimate from the experiment would be expected fall within the range of the $\hat{\sigma}_{\alpha_i}$ that arises from the permutation analysis. Figure A2 shows that this is not the case.

Distribution of fixed−effect standard deviation
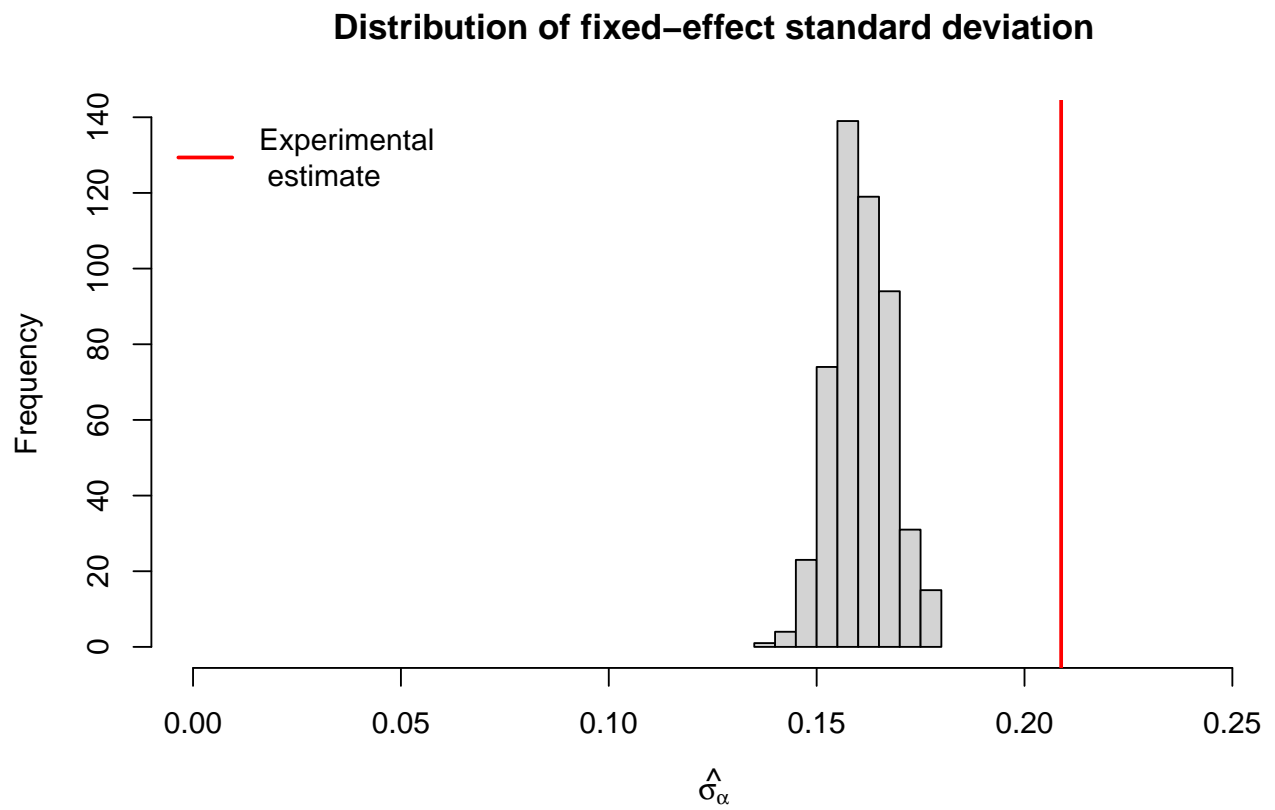
Figure A2: Distribution of $\hat{\sigma}_\alpha$
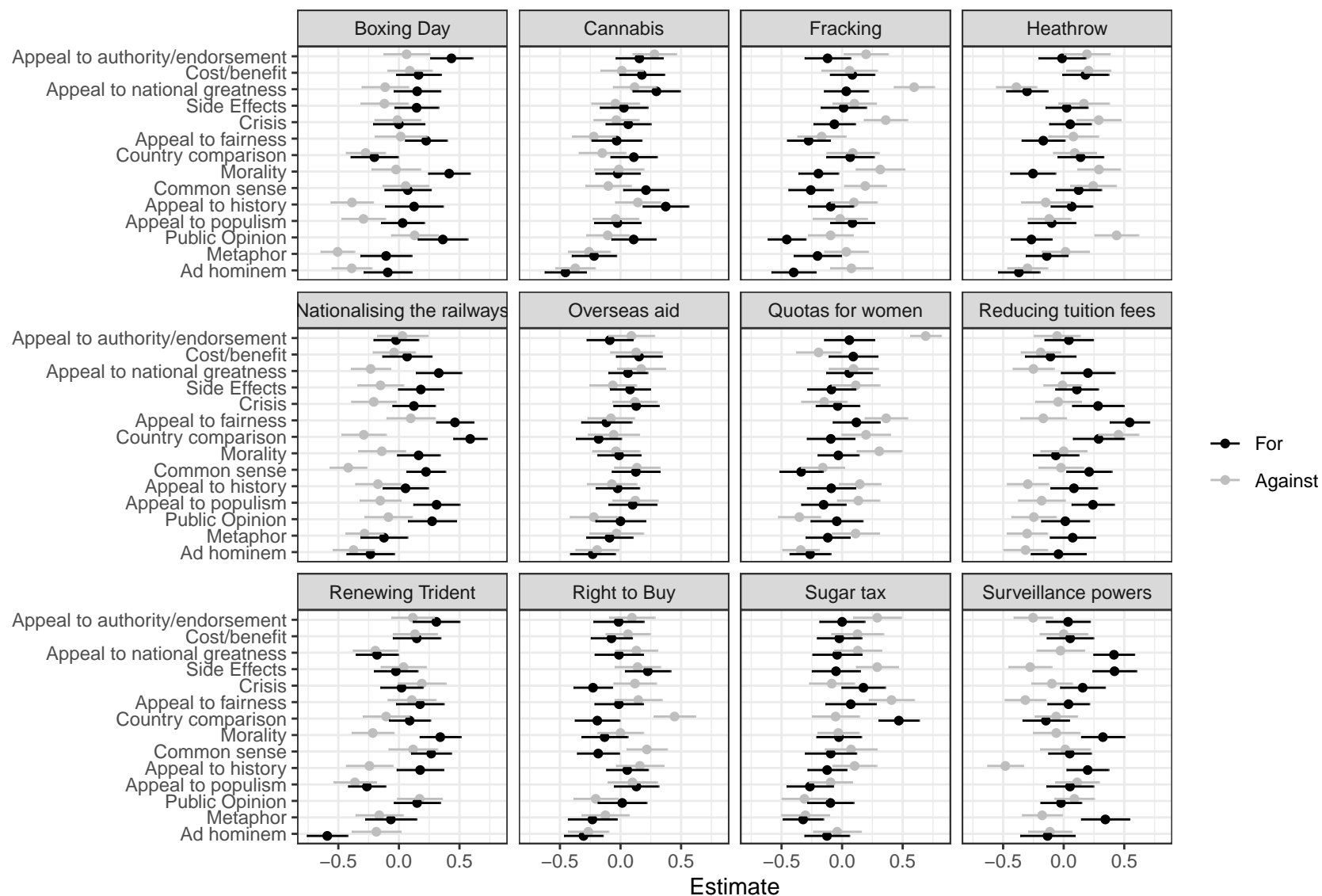
Figure A3: Argument strength, by rhetorical element
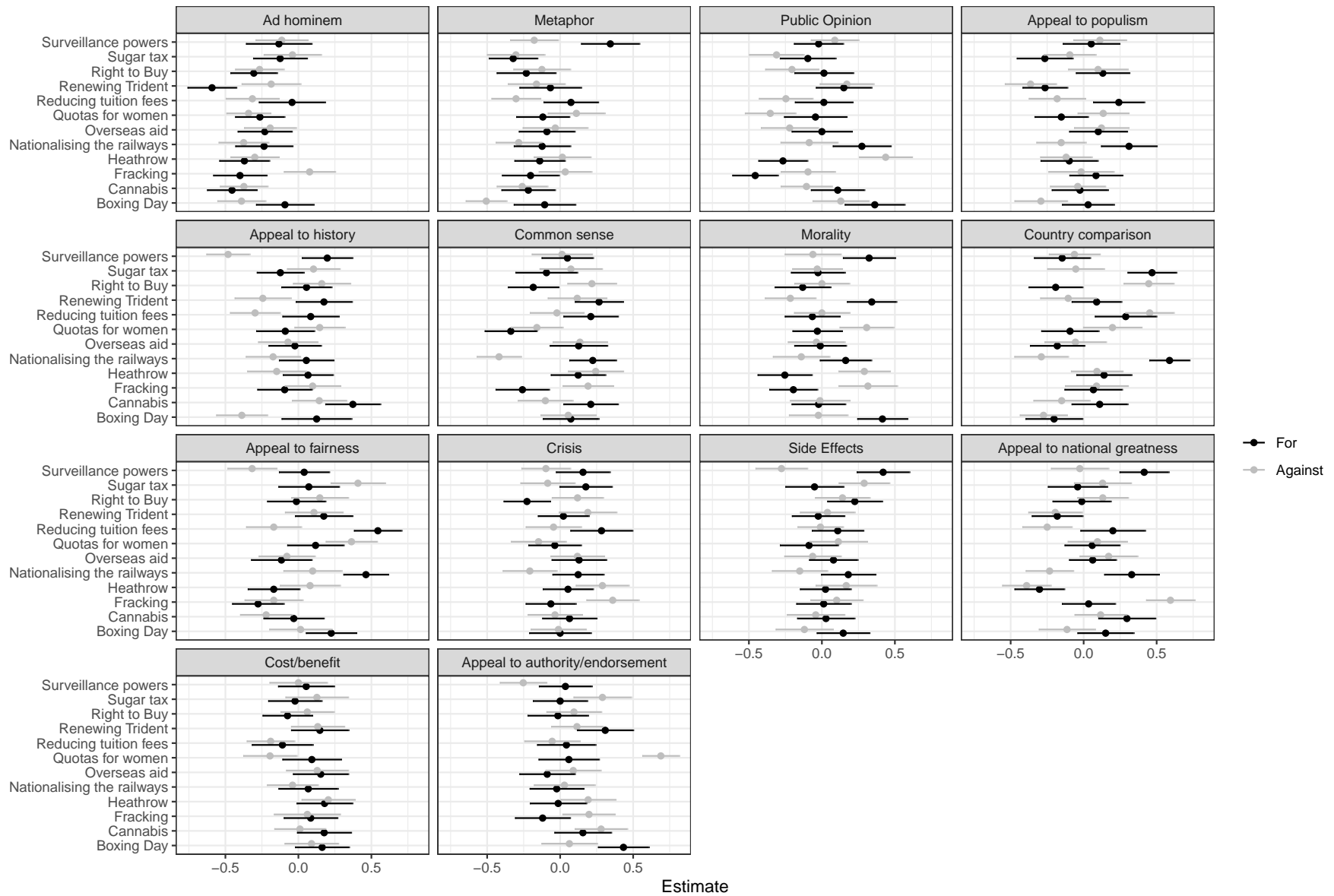
Figure A4: Argument strength, by policy area
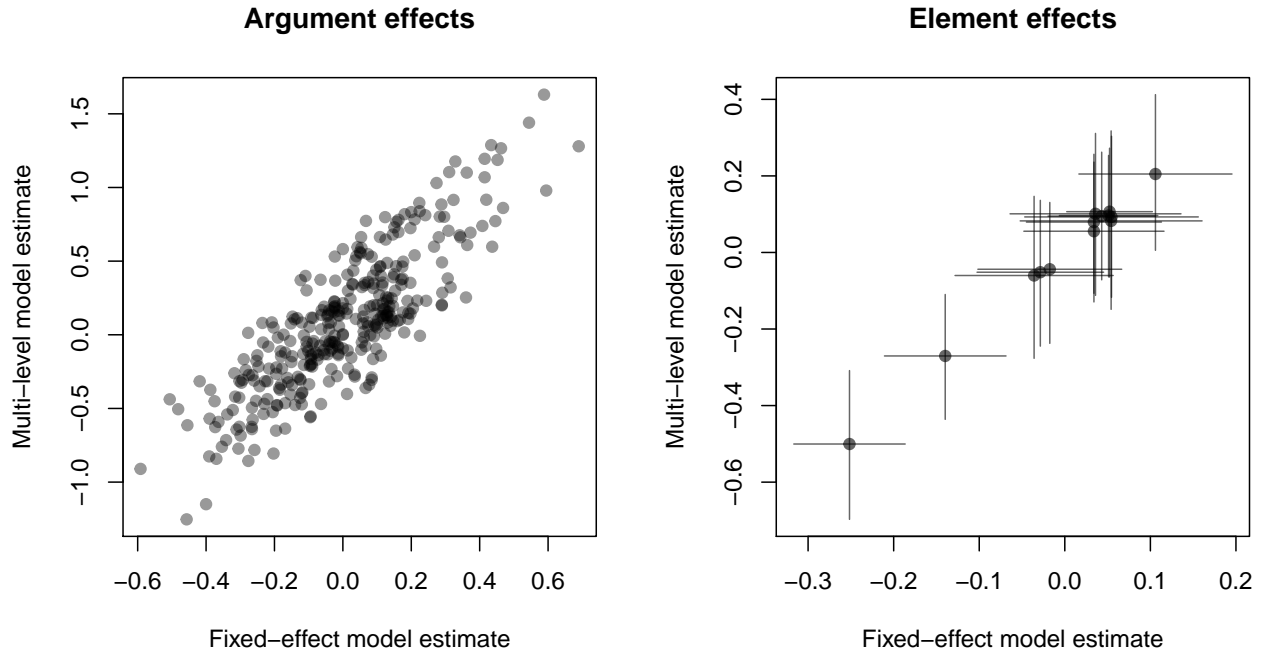
Figure A5: Comparison of element and argument strength estimates

Element-level estimates

In the right-hand panel of figure A5 we compare the estimates of the persuasive power of each of our 14 rhetorical elements to estimates derived from a simpler fixed-effect model. We again stack the data such that our outcome, $Y_{i(j,r)}$, is 1 (-1) if respondent r rates argument j as more (less) persuasive in a given comparison, and 0 if the respondent rates both arguments in the comparison as "About the same". We then model this outcome using a regression of the form:

$$Y_{i(j,r)} = \mu_e(j) + \epsilon_{i(j,r)} \tag{A2}$$

where μ_e are fixed-effects for each of our rhetorical elements. As above, we estimate the model using OLS, omitting the intercept, and here clustering standard errors at the argument level.

Figure A5 shows that we get a very similar distribution of estimated element strengths using this approach. As we document using the main model specification, only metaphors and *ad hominem* attacks appear to be substantively less persuasive than arguments that employ other elements, with only small differences between the other element types. The correlation between

the point estimates from our multi-level model and the model described in equation A2 is 0.997.

We take this as evidence that our central results are robust to different modelling strategies.

Additional Model Parameter Estimates

The plots in this section show additional details for the main model in the paper. Figure A6 shows the values of the δ parameters from the model, which describe the extent to which argument strength depends on which side of an issue is taken by the argument. Figure A7 shows the strength of evidence regarding the relative average strength of all pairwise comparisons of rhetorical element types.
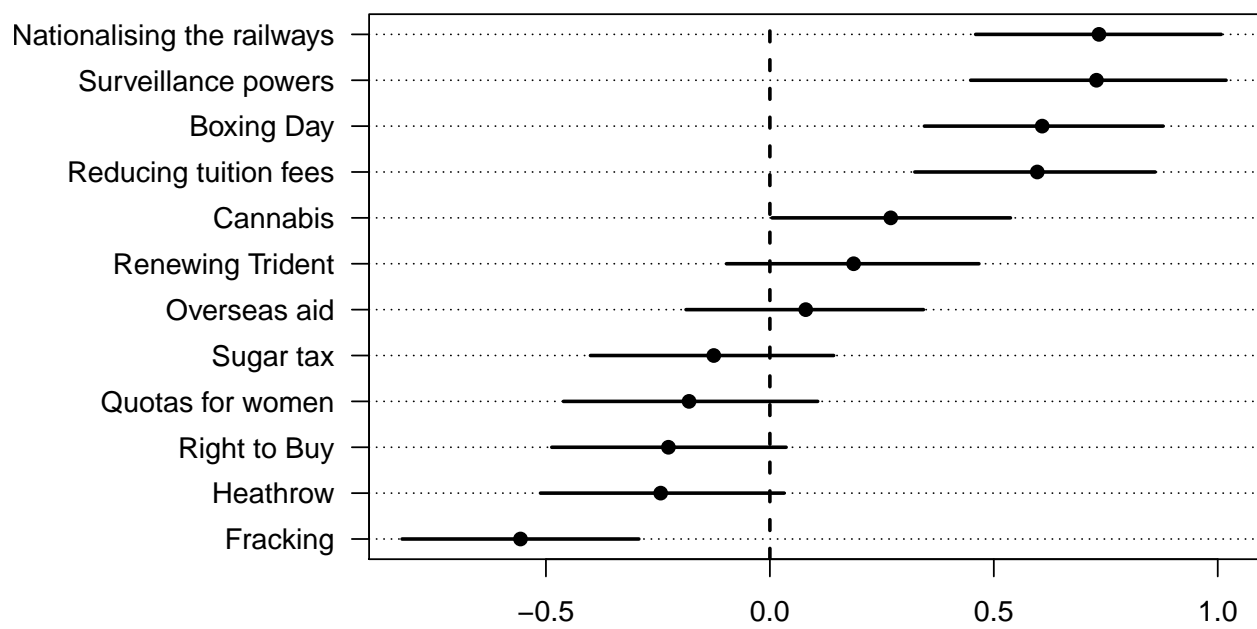
Figure A6: Policy effects

	authority	cost/benefit	national greatness	side effects	crisis	fairness	country comparison	morality	common sense	history	populism	public opinion	metaphor	ad hominem
authority		0.83	0.8	0.86	0.86	0.81	0.83	0.87	0.91	0.98	0.99	0.98	1	1
cost/benefit	0.17		0.52	0.55	0.55	0.55	0.58	0.62	0.72	0.94	0.95	0.93	1	1
national greatness	0.2	0.48		0.52	0.54	0.53	0.56	0.58	0.66	0.89	0.9	0.89	1	1
side effects	0.14	0.45	0.48		0.51	0.51	0.55	0.59	0.69	0.95	0.95	0.93	1	1
crisis	0.14	0.45	0.46	0.49		0.51	0.54	0.57	0.67	0.94	0.94	0.92	1	1
fairness	0.19	0.45	0.47	0.49	0.49		0.52	0.55	0.64	0.89	0.89	0.88	1	1
country comparison	0.17	0.42	0.44	0.45	0.46	0.48		0.51	0.59	0.85	0.86	0.85	1	1
morality	0.13	0.38	0.42	0.41	0.43	0.45	0.49		0.6	0.89	0.9	0.89	1	1
common sense	0.09	0.28	0.34	0.31	0.33	0.36	0.41	0.4		0.85	0.85	0.85	1	1
history	0.02	0.06	0.11	0.05	0.06	0.11	0.15	0.11	0.15		0.53	0.56	0.99	1
populism	0.01	0.05	0.1	0.05	0.06	0.11	0.14	0.1	0.15	0.47		0.53	0.99	1
public opinion	0.02	0.07	0.11	0.07	0.08	0.12	0.15	0.11	0.15	0.44	0.47		0.97	1
metaphor	0	0	0	0	0	0	0	0	0	0.01	0.01	0.03		0.99
ad hominem	0	0	0	0	0	0	0	0	0	0	0	0	0.01	

Figure A7: The figure shows the posterior probability that the average persuasive effect of any rhetorical element type (row) is greater than any other rhetorical element type (column)

Differential Persuasiveness by Respondent Characteristics

Given the modest average differences in the strength of rhetorical elements described above, our experiment is not sufficiently well powered to detect differences in the effectiveness of rhetorical *type* by respondent demographics. Nonetheless, the degree of heterogeneity at the argument-level enables us to assess the degree to which different groups of respondents are persuaded by the same arguments in general, even though we are very far from having enough data to do so at the level of individual arguments.

In order to assess the extent to which different groups of people perceive the relative persuasiveness of different arguments differently, we use a "correlated strength" model to assess the possibility that arguments might have different relative strengths for different groups of respondents. These strengths for respondent groups $g \in 1, \ldots, G$ are assumed to follow a multivariate normal distribution with mean zero and covariance matrix Σ with diagonal elements σ_g^2 and off-diagonal elements $\sigma_g \sigma_{g'} \rho_{gg'}$. The correlations ρ are our primary interest, as these tell us whether the relative strengths of the arguments, across our entire experiment, tend to be very similar for pair of groups g and g' (large $\rho_{gg'}$, near 1), whether the arguments that are persuasive to one group are uncorrelated with those that are persuasive to the other (small $\rho_{gg'}$, near 0), or whether the groups systematically disagree about which arguments are strong (negative $\rho_{gg'}$). Where g is a group of respondents, the correlated strength model is specified as:

$$\alpha_{jg} = \delta_{p(j)} + \nu_{jg}$$
$$\nu_{jg} \sim MVN(0, \Sigma) \tag{A3}$$

The terms relating to the average argument quality of element types and the control variables introduced in the last section are the same in expectation for respondents in different groups due to the experimental randomisation, and so we omit these model terms to simplify interpretation here.

When we apply this model to the data from our experiment, we estimate that the relative persuasiveness of the arguments for 2017 Conservative voters had a correlation with the relative persuasiveness for Labour voters of 0.46. Comparing 2016 Leave voters with Remain voters,

the corresponding correlation is 0.79. Among demographic variables, we see the lowest correlations when we compare by age, 0.66 comparing 18-34s with 55+s (0.85 comparing 18-34s with 35-54s, 0.87 comparing 35-54s with 35-54s). We see relatively high correlations when we compare men and women (0.88), those with and without degrees (0.94), and those with high versus low attention to politics (0.97).

These correlations are all positive, which is consistent with other work that finds homogeneity in treatment effect sizes across different groups of survey respondents (Coppock, Leeper and Mullinix, 2018). There is more disagreement between Conservative and Labour voters about which arguments are relatively persuasive than as a function of any other variable we test above, followed by age and referendum vote, with mostly negligible differences by gender, education and attention to politics. At least some of these are non-obvious: one might have expected education and attention to politics to be more important in shaping which arguments people find convincing. It is also perhaps surprising that referendum vote structures views of persuasiveness relatively weakly compared to general election vote and age.

Same-side vs Opposite-side Argument Comparisons

As we have already discussed, a novel feature of our experiment is that it allows us to assess relative argument strength using either pairwise comparisons of arguments on the same-side of an issue or comparisons of arguments on opposite-sides. Is it the case that the same arguments perform relatively well in both types of comparisons?

We again use the "correlated strength" model to assess whether arguments have the same relative strengths in same-side and opposite side comparisons. Σ is once again a 2x2 covariance matrix describing the distribution of argument strengths in opposite-side and same-side comparisons, with diagonal elements σ_1^2 and σ_2^2 and off-diagonal elements $\rho\sigma_1\sigma_2$. The magnitudes of σ_1 and σ_2 indicate the variation in argument strengths in opposite-side and same-side comparisons, respectively, and ρ indicates the correlation between the strength of arguments in these two types of comparisons.

When we fit this model, we recover the following mean posterior and 95% central interval estimates for the parameters described above: $\rho = 0.82$ (0.69-0.93), $\sigma_1 = 0.33$ (0.28-0.38), and

$\sigma_2 = 0.47$ (0.42-0.52). The correlation ρ is large, indicating that we seem to be recovering nearly the same information about the relative strength of arguments from the same-side and opposite-side comparisons. The fact that σ_1 is smaller than σ_2 means that variation in measured argument strengths is less predictive of respondent choices in the opposite side comparisons than in the same side comparisons. This is what we would expect if opposite-side comparisons in part reflect respondents' policy views.

Collectively these results indicate that we can get nearly the same information from same-side comparisons as from opposite-side comparisons, which was not obvious ex ante. This is useful to know in light of the fact that respondent preferences for arguments on one side of the argument versus the other do seem to suppress the signal we get about argument strength from opposite-side comparisons. The combination of these two results speaks to an important tradeoff in designing experiments to test the strength of arguments. In many contexts, one is ultimately interested in the performance of arguments against arguments on the other side of the issue. However, these results indicate that, on a per response/respondent basis, you can get a comparably strong signal about the performance of arguments against opposite side counter-arguments by instead testing them against other arguments on the same side. The differences in relative performance in the two contexts are are modest in magnitude and the signal from respondents is substantially stronger in the same side comparisons.

Rhetorical Fit to Specific Issues

It is possible that certain rhetorical arguments are a better fit to certain issues. If this were the case, we might expect to see positive correlations between the strength of the argument for and the argument against a given policy using the same rhetorical element. Using the same correlated strength model used in the preceding sections of this appendix, we can assess whether there is such a correlation.

When we fit this model, we recover the following mean posterior and 95% central interval estimates for the correlation parameter: $\rho = 0.44$ (0.25-0.60). Thus, we find evidence that there is a general tendency for rhetorical types to fit certain issues better than others: if a rhetorical type is relatively effective in making the case in favour of a policy, it also tends to be relatively

effective in making the case against that policy. This correlation is moderate in magnitude, but there is strong evidence that it is positive in the set of arguments that we constructed.

Multilevel Model for Validation Experiment

In the main text of the paper, we analyse the validation experiment using the simple differences in means between the "strong in favour" and "strong against" conditions for each issue, where responses endorsing "For" are coded $Y = 1$, "Not sure" $Y = 0.5$, and "Against" $Y = 0$. Here, we additionally present results from a multilevel model. This model takes the form:

$$
\begin{aligned}
Y_{j,j'} &= \theta_p + \alpha_j - \alpha'_j + \epsilon \\
\alpha_j &= \delta_{p(j)} * \mathsf{Strong}_j + \\
&\quad \gamma_{Words} * \#\,\mathsf{Words}_j + \\
&\quad \gamma_{Arguments} * \#\,\mathsf{Arguments}_j + \nu_j \\
\nu_j &\sim N(0, \sigma) \\
\delta_p &\sim N(\mu_\delta, \sigma_\delta)
\end{aligned}
$$

where θ_p is the baseline popularity of the "For" side of each policy area p, $\delta_{p(j)}$ is a parameter capturing the effect of using strong (as opposed to weak) arguments on persuasive power for policy area p, γ_{Words} is the effect of the number of words, and $\gamma_{Arguments}$ captures whether paragraphs with three rather than two arguments are more persuasive. In this analysis, our main quantities of interest are $\delta_{p(j)}$—which measure the average treatment effect of going from weak arguments in favour and strong arguments against to strong arguments in favour and weak arguments against for a particular issue area—and μ_δ—which measures the average of these average treatment effects across the set of issues in the experiment. The way that the $\delta_{p(j)}$ and μ_δ are defined in the multilevel model as moving from a weak to a strong argument on one side means that they correspond to half of the experimental difference in means, which correspond to moving from weak to strong on one side and strong to weak on the other.

The right-hand panel of figure A8 presents median posterior estimates and 95% intervals for the key parameters from the multilevel model described in equation A4. Neither the number
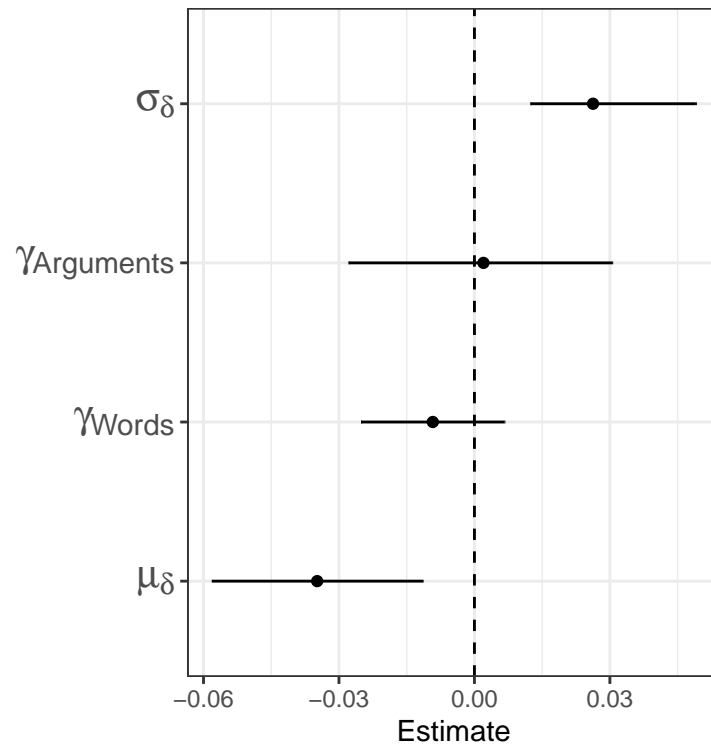
Figure A8: Experiment two multilevel model results

of words nor the number of arguments presented significantly predict responses. We can rule out the possibility that there is no variation in the average treatment effects across issues (the interval estimate for σ_δ starts well above 0). Finally, our estimate of μ_δ, the average effect of going from a weak to a strong argument, is -0.035, which translates to an average treatment effect of -7 percentage points on the scale of the simple difference in means discussed previously.

MCMC diagnostics

Figure A9: MCMC chains for $\mu_{e(j)}$ parameters in equation 3

Figure A10: MCMC chains for $\mu_{e(j)}$ and γ_k parameters in equation 4

Arguments

Table A1: Treatment texts – Building a third runway at Heathrow

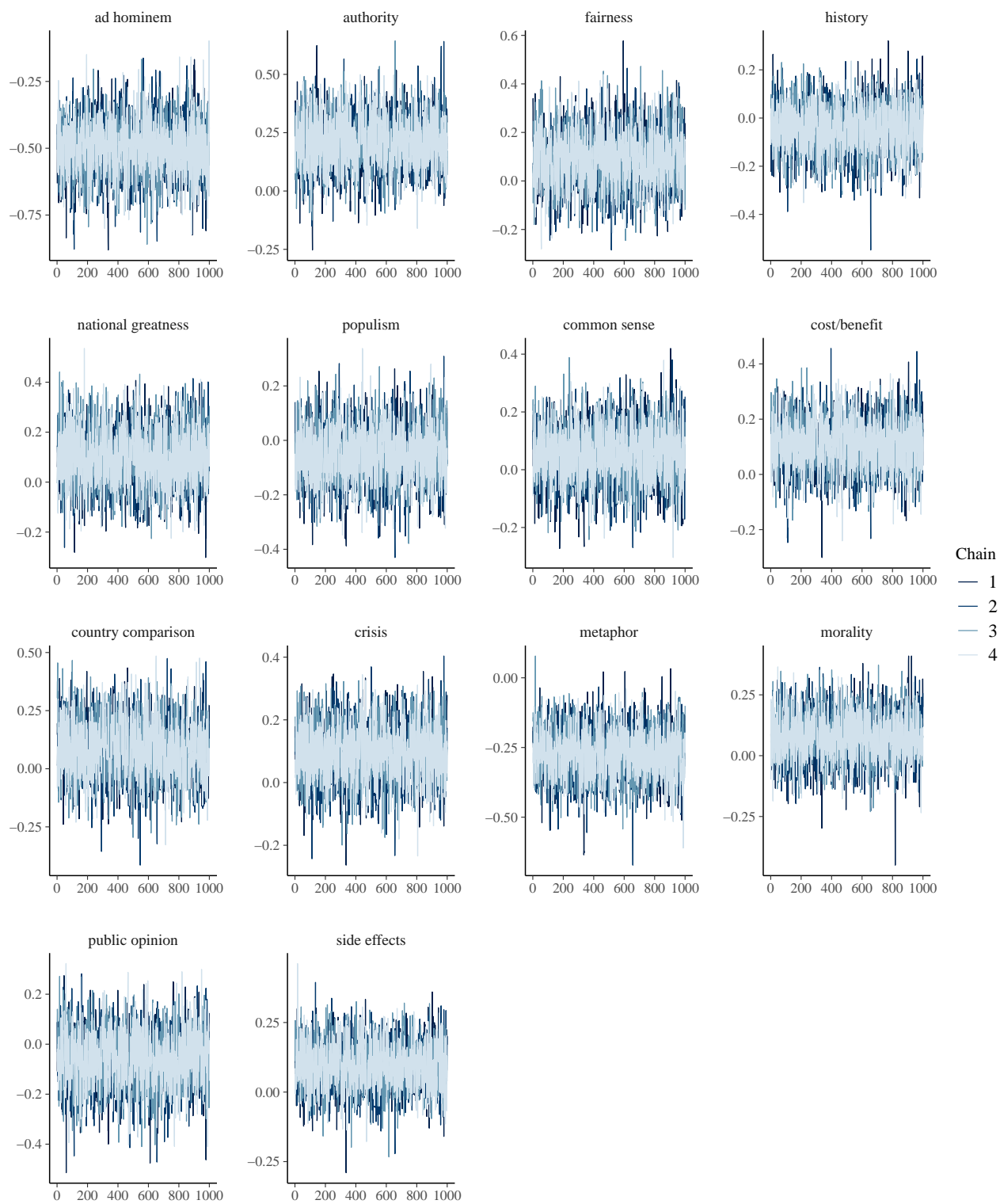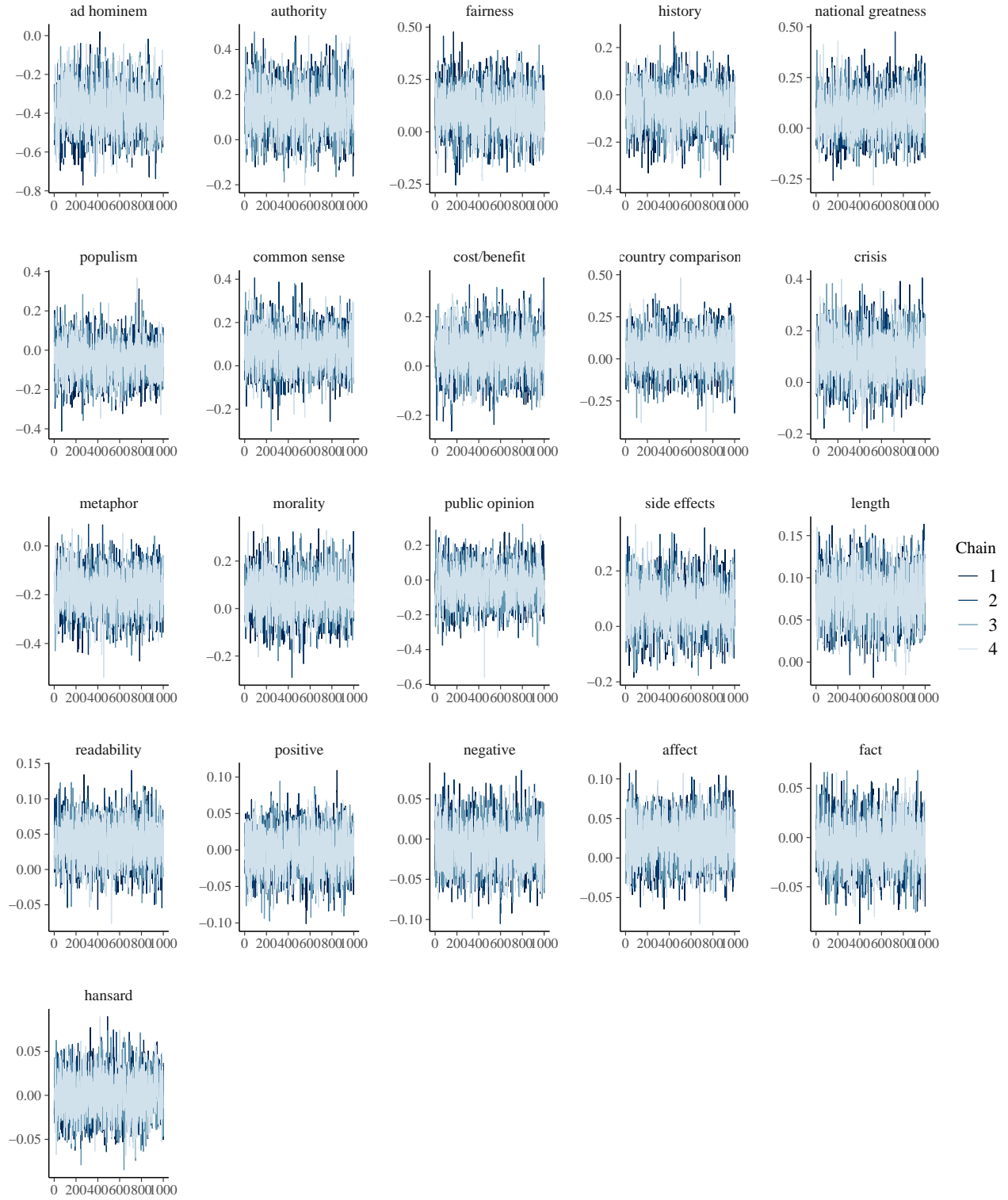| Element	For/Against	Text
Ad hominem	For	The arguments against the expansion at Heathrow are being made in bad faith by people with narrow and selfish interests.
	Against	Building a third runway would be giving a blank cheque to the foreign-owned multinational company that runs Heathrow.
Appeal to authority	For	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".
	Against	Research from Greenpeace suggests that Heathrow expansion would increase greenhouse-gas emissions to levels that would be irreconcilable with internationally agreed levels.
Appeal to fairness	For	The decision to expand Heathrow is the result of a fair and transparent process which weighed all the relevant concerns.
	Against	The Heathrow expansion is another example of London and the south-east getting more than their fair share of investment, spending, jobs and benefits.
Appeal to history	For	We must learn from the lessons of history: failure to invest in transport infrastructure projects like the Heathrow expansion leads to increasing delays and increasing prices.
	Against	History show us that most large infrastructure projects do not lead to significant economic growth, which suggests that the expansion of Heathrow will fail to pay for itself.
Appeal to national greatness	For	Only by expanding Heathrow will Great Britain get the great international airport it deserves.
	Against	Great nations don't waste money on vanity projects, and the expansion of Heathrow would be nothing more than a project of national vanity.
Appeal to populism	For	Air travel should not just be for elites: expanding Heathrow will make flights cheaper and more accessible for ordinary working people.
	Against	Expanding Heathrow will enrich a private foreign-owned business at the expense of higher fares for ordinary passengers.
Common sense	For	It is just common sense that an airport as congested as Heathrow should be expanded.
	Against	Given the obvious problems of expanding Heathrow, the common sense solution is to build additional capacity elsewhere.
Cost/benefit	For	Expansion at Heathrow will bring real benefits across the country, including a boost of up to 74 billion to passengers and the wider economy, and these will easily surpass the costs of expansion.
	Against	The unnecessarily large costs of expansion at Heathrow will end up falling on taxpayers and airline passengers, and will outweigh any potential benefits.
Country comparison	For	Major European international airports have more runways than Heathrow, putting the UK at a competitive disadvantage.
	Against	Expansion is not necessary, as Heathrow already has more international flights each week than either Charles de Gaulle or Frankfurt, both of which have more runways than Heathrow.
Crisis	For	Heathrow has a capacity crisis: without a third runway, delays will increase and more flights will be cancelled.
	Against	The crisis of noise and air pollution around Heathrow will only be made worse by the addition of a third runway.
Metaphor	For	Heathrow does not need expanding, as it is already the beating heart of the UK aviation network.
	Against	Heathrow is the clogged artery of the UK aviation network, and expansion is the best possible treatment.
	For	It is wrong to let local interests get in the way of the national interest in having an internationally competitive airport.

Element	For/Against	Text
Morality	Against	As air travel is one of the largest contributers to the world's carbon emissions, we have a moral responsibility to reduce air travel, not to increase it by building bigger airports.
Public Opinion	For	A 2014 poll showed that many more people were in favour of a third runway at Heathrow than opposed it.
	Against	Recent polling shows that more people want to expand other London airports, rather than expanding Heathrow.
Side Effects	For	In addition to the immediate benefits of Heathrow expansion for passangers, it will have the side-effect of benefitting businesses and individuals right across the country by bringing additional growth to the UK economy.
	Against	An unintended consequence of expansion at Heathrow would be to add further vehicle traffic to an already overburdened motorway system around London.

Table A2: Treatment texts – Closing large retail stores on Boxing Day

Element	For/Against	Text
Ad hominem	For	The people who support Boxing Day opening are the people who get rich by exploiting their workers.
	Against	The people who oppose Boxing Day sales are the sort of people who can afford to pay full price for everything.
Appeal to authority	For	The Union of Shop, Distributive and Allied Workers says that ever increasing opening hours have a big impact on workers and their families during the Christmas period.
	Against	The British Retail Consortium say that Boxing Day trading is increasingly important to the profitability of many businesses.
Appeal to fairness	For	Boxing Day opening is particularly unfair to retail workers who have to work when others get the day off.
	Against	It is unfair for the Government to tell businesses to close their shops on one of the busiest days of the year.
Appeal to history	For	Historically, Boxing Day closure was normal because it was a bank holiday, and no business thought of doing anything other than closing.
	Against	Historically, people across many sectors worked on Boxing day.
Appeal to national greatness	For	A great country like ours should put families not shopping at the centre of its holiday tradition.
	Against	Boxing day sales are part of our great national Christmas holiday tradition.
Appeal to populism	For	Boxing Day opening is only good for business elites, not for ordinary people.
	Against	Most people like Boxing Day sales, it is typical elitism to try to ban them.
Common sense	For	It is just common sense that large stores, which are open almost every day, should be closed on a public holiday.
	Against	The current law provides a common sense balance between the interests of employers and workers with regard to Boxing Day trading.
Cost/benefit	For	The benefits to the 365,000 people in the UK retail industry who work on Boxing Day would far outweigh any costs to retailers.
	Against	Banning high street outlets from opening the day after Christmas could cost retailers the 3.7 billion that was spent on Boxing Day last year.
Country comparison	For	No other country in the world has anything comparable to our tradition of Boxing Day sales and their companies are fine without it.
	Against	In the vast majority of countries, people do not have the day off work on Boxing Day.

Table A2: Treatment texts – Closing large retail stores on Boxing Day *(continued)*

Element	For/Against	Text
Crisis	For	Boxing Day shopping is indicative of the compulsive consumerism that has left family life in crisis in this country.
	Against	Retail is in crisis and banning Boxing Day trading would only make matters worse.
Metaphor	For	We have made Scrooge's policy of forcing workers back to their jobs on the day after Christmas into our national policy.
	Against	Only a real Scrooge would ban shopping at Christmas time.
Morality	For	It is wrong to create a society where people have to work on holidays instead of spending time with their families.
	Against	It would be wrong for the government to interfere with Boxing Day trading by telling businesses how to run their shops and serve their customers.
Public Opinion	For	In a recent survey, 92 per cent of workers said that they did not want to have to work on Boxing Day.
	Against	People vote with their feet, and huge numbers of the British public demonstrate each year how popular Boxing Day shopping is.
Side Effects	For	An unintended consequence of shops being open on Boxing Day is that workers in a whole range of related services, such as waste collection, emergency services and transport, also have to work on a public holiday.
	Against	An unintend consequence of banning Boxing Day opening would be to accelerate the decline of our high streets.

Table A3: Treatment texts – Extending the Right to Buy

Element	For/Against	Text
Ad hominem	For	People who oppose the right to buy are hypocrites who already own their own homes.
	Against	Right to buy is a scheme designed by people who want to eliminate public housing.
Appeal to authority	For	The National Housing Federation has argued that the right to buy will be good for residents, housing associations, and for the country's housing supply.
	Against	The Chartered Institute of Housing has described extending the right to buy as a "fire sale" policy which will lead to a significant decline in affordable housing.
Appeal to fairness	For	We should extend the right to buy because it is unfair that council tenants can buy their homes but housing association tenants cannot.
	Against	Extending the right to buy would be unfair, as the lack of social housing means that people can no longer live in the communities where they were born.
Appeal to history	For	The introduction of the right to buy in the 1980s delivered one of the biggest transfers of wealth to working people of any policy in British history.
	Against	We should learn lessons from recent history, as the introduction of the right to buy in the 1980s decimated the social housing stock of this country.
Appeal to national greatness	For	Owning one's home is a central part of the Great British dream, and extending the right to buy will help more people to realise that dream.
	Against	Communities are at the centre of our great nation and we need a radical, ambitious, compassionate housing policy to strengthen them.
	For	Extending the right to buy will make home ownership possible for more ordinary people, not just for elites who can pay increasingly outrageous prices.

A22

Table A3: Treatment texts – Extending the Right to Buy *(continued)*

Element	For/Against	Text
Appeal to populism	Against	Extending the right to buy will mean that prime properties will be flogged off to speculators and to buy-to-let landlords, not to ordinary families who are struggling to buy for the first time.
Common sense	For	Introducting the right to buy for housing associations is a common sense approach to housing policy.
	Against	It is common sense that if you force housing associations to sell off their properties, it will undermine those associations and make our housing problems worse.
Cost/benefit	For	The potential benefits of extending the right to buy to the housing association sector are huge, as it would give 1.3 million tenants the right to become homeowners.
	Against	Regardless of any benefits, extending the right-to-buy is too costly: it will add billions to the government's housing budget.
Country comparison	For	The existing stock of social housing in the UK is far larger than in either the US or in France and so we are well positioned to extend the right-to-buy.
	Against	In Scotland, where the government ended right-to-buy, waiting lists for council housing have shortened dramatically.
Crisis	For	Claims of a housing crisis in the UK are overblown; it would be a disaster for many aspirational people to be denied the opportunity to own their own homes.
	Against	The UK is in the midst of a chronic housing crisis, and extending the right to buy will only make it worse.
Metaphor	For	An Englishman's home is his castle, and extending the right-to-buy will give more people the opportunity to own a castle of their own.
	Against	Extending privitisation to housing association properties will rip the heart out of social housing in the UK.
Morality	For	It is our moral duty to ensure that everyone has the oppurtunity to own their own home.
	Against	It is immoral to further deplete the social housing stock through an extension of the right to buy policy.
Public Opinion	For	Home ownership is an aspiration for 86 per cent of people in this country and extending the right to buy will make this attainable for more of them.
	Against	The public greatly prefers public renting to private renting, yet extending the right to buy would push more people into private rentals.
Side Effects	For	Extending the right to buy might have positive side effects, as people who own their own property have more incentive to look after it and the neighbourhood around it.
	Against	One side effect of extending the right to buy is that it will be much more difficult for housing associations to function.

Table A4: Treatment texts – Extension of surveillance powers in the UK

Element	For/Against	Text
Ad hominem	For	The people who are fighting new surveillance powers put no value on your safety.
	Against	Law enforcement agencies always want more information, they put zero value on your privacy.
Appeal to authority	For	The National Crime Agency has made it clear that it is essential that our security services have more power to intercept electronic communications.
	Against	The National Council for Civil Liberties has argued that there is "no operational case" for the extension of surveillance powers.
	For	Extending surveillance powers is the fairest way to protect all people in our country.

Table A4: Treatment texts – Extension of surveillance powers in the UK *(continued)*

Element	For/Against	Text
Appeal to fairness	Against	Excessive survellance always ends up being unfair to law abiding people, no matter the promises made.
Appeal to history	For	Historically, this country has repeatedly adapted how it engages in policing as communication technology has changed.
	Against	The historical abuses of state surveillance powers are too many to list.
Appeal to national greatness	For	The quiet heroism of our security services should be a source of great national pride, and we must provide the tools they need to maintain their excellence.
	Against	We have a great tradition in Britain of protecting our liberty and our privacy from unnecessary state surveillance.
Appeal to populism	For	We must protect ordinary people rather than cater to elite sensibilities about policy surveillance.
	Against	We must ensure that we have safeguards that protect ordinary people from surveillance, intrusion and abuse by unaccountable elites.
Common sense	For	The current law is clearly out of date, and so it is common sense to introduce new surveillance powers for new technologies.
	Against	It is common sense to be concerned about ever increasing surveillance powers that reach further and further into the lives of everyone in this country.
Cost/benefit	For	The small risk that surveillance powers will be inappropriately used is far outwieghed by the benefits they bring.
	Against	Law enforcement agents like to claim the benefits of extending their surveillance powers will be large, but they never talk about the costs of doing so.
Country comparison	For	In Germany, where the security services do not have necessary surveillance powers, successful investigations of terrorist-related individuals are much rarer than in the UK.
	Against	We should be reluctant to extend police surveillance powers when judicial oversight of the use of these powers is weaker in the UK than it is in the USA, Australia, Canada, or New Zealand.
Crisis	For	The crisis of recent attacks around the globe show the need for increased surveillance powers.
	Against	The reason that supporters of extending surveillance powers pretend there is a security crisis is that they know people will not support these measures without one.
Metaphor	For	Without new powers, our law enforcement agencies will become increasingly unable to cope with the spider's web of criminal activity that is organised online.
	Against	Extending surveillance powers allows the tentacles of government to wind their way further and further into our private lives.
Morality	For	The primary moral duty of any government is to keep its people safe, and that is what increased surveillance powers will achieve.
	Against	It is immoral to engage in the mass collection of private data from people who have not been accused of doing anything wrong.
Public Opinion	For	All surveys of public opinion suggest that people have a very high level of confidence in our intelligence and security services, and so extending surveillance powers will not be of great concern to them.
	Against	The public do not have confidence in the system of checks and balances that currently regulates our surveillance system, let alone for a new set of investigatory powers.
Side Effects	For	A side-effect of increased surveillance powers is to make it more difficult for criminal and terrorist organisations to recruit.
	Against	One unattractive side-effect of the extension of surveillance powers is the chilling effect it would have on journalists, who need to be able to gather information without the fear of their sources being exposed.

Element	For/Against	Text
Ad hominem	For	Anti-fracking environmental messages are often promoted by Russia with the goal of maintaining other countries' dependence on Russian gas and oil supplies.
	Against	Fracking companies are interested in their own short-term profits, this push for drilling is nothing more than selfishness.
Appeal to authority	For	The government's Chief Scientific Advisor has said that fracking could dramatically increase accessible UK natural gas reserves.
	Against	The Committee on Climate Change, the government's official advisers, has said that shale gas production through fracking will break the UK's climate change targets.
Appeal to fairness	For	It is unfair to make people pay more to heat their homes because of misplaced concerns about fracking.
	Against	Fracking is unfair to those residents who live closest to the drilling sites.
Appeal to history	For	The UK has a long history of effective regulation of successful and safe onshore and offshore gas extraction, and there is no reason for fracking to be any different.
	Against	History shows us that government regulation does not prevent environmental disasters and fracking creates a high risk of accidents.
Appeal to national greatness	For	Fracking has the potential to help make the UK a world leader in new energy technologies, and to make it self-sufficient in energy production.
	Against	The UK should be leading the world by adopting clean and renewable energy sources, not investing further in fossil fuels such as those produced by fracking.
Appeal to populism	For	Fracking will enable us to reduce the cost of fuel, which may not be much concern to elites, but matters a great deal to ordinary people who are struggling to get by.
	Against	Fracking will produce windfall profits for energy multinationals, but little benefit for ordinary people.
Common sense	For	It is common sense that we should use fracking to take advantage of the natural gas resources that we have.
	Against	It is common sense that as we are trying to address climate change we should not be using fracking to increase our use of fossil fuels.
Cost/benefit	For	The benefits of fracking include both lower gas prices and jobs in communities that desparately need them.
	Against	The benefits that could results from fracking are too uncertain to justify the costs of extracting gas in this way in the UK.
Country comparison	For	In the US, the development of new fracking technologies has helped to increase domestic energy production, reduced carbon emissions and improved the security of energy supply.
	Against	In the US, fracking has caused increased earthquakes and water supplies so tainted with gas that kitchen taps can sustain a flame.
Crisis	For	The climate change crisis can be most effectively addressed if we use all the tools at our disposal to reduce carbon emissions: gas from fracking is better for the environment than coal or oil.
	Against	The climate change crisis cannot be effectively addressed through fracking: we will just be switching between different non-renewable fossil fuel energy sources.
Metaphor	For	Fracking can provide a bridge to a greener future.
	Against	Fracking is another scorched earth resource extraction technology.
Morality	For	It is our moral duty to take responsibility for generating our own energy supplies, and fracking can help us to achieve this.
	Against	It is our moral duty to protect the environment from a risky fracking scheme with unclear benefits.

Element	For/Against	Text
Public Opinion	For	The tiny group of people who protest about fracking are not at all representative of public opinion.
	Against	A recent survey showed that only 19 per cent of people support fracking.
Side Effects	For	One side-effect of fracking would be to decrease the usage of oil and coal, thereby reducing pollution and greenhouse gas emissions.
	Against	One side-effect of exploratory fracking seems to be that it can cause small earthquakes.

Table A6: Treatment texts – Nationalisation of the railways in the UK

Element	For/Against	Text
Ad hominem	For	Currently, train fare increases go straight into the pockets of greedy franchise owners.
	Against	Greedy unions think that it will be easier for them to get large wage increases from a nationalised railway system.
Appeal to authority	For	A recent report by the Institute for Public Policy Research shows that privitisation has not delivered on its promises, as it still takes longer to get from Liverpool to Hull than it does to get from London to Paris.
	Against	Experts from Edinburgh Napier University suggest that nationalisation is unlikely to solve existing problems facing the UK's railways.
Appeal to fairness	For	In recent years, fares have increased three times faster than wages, which is unfair to people who can no longer afford them.
	Against	It is not fair to ask people who do not use the railways to pay for them through the increased taxes that would be required for nationalisation.
Appeal to history	For	The current franchise system is a mess that emerged from the hurried privatisation in the 1990s.
	Against	The current franchise system is enabling a larger investment in railways infrastructure than at any time since the Victorian era.
Appeal to national greatness	For	Great Britain deserves to have a great railway system, not a collection of mediocre franchises.
	Against	It was private investment that created a great railway system in this country, and it was nationalisation that did so much to diminish it.
Appeal to populism	For	Normal people are being priced out of the railways; nationalisation would make sure that the railways are accessible for all.
	Against	Nationalisation will just hand more power to London elites and there will be even less attention paid to the transport needs of ordinary people around the country.
Common sense	For	If private sector rail franchises repeatedly fail, it is only common sense to take them into public ownership permanently.
	Against	Private sector rail franchises are a common sense way of administering the railway network.
Cost/benefit	For	The current system is one where franchise owners get most of the benefits and the taxpayer is left with most of the cost.
	Against	Nationalisation is unlikely to improve the quality of service for the public, and it would inevitably be disruptive, costly and time-consuming.
Country comparison	For	Rail fares are a great deal more expensive in the UK than in countries like Germany, where the government owns and controls significantly more of the railway infrastructure.
	Against	Under the current private system, use of the rail network has grown faster than in most European countries.
	For	The current franchise system has been responsible for a series of transport crises throughout the country.

Element	For/Against	Text
Crisis	Against	The railways are not in the crisis some would have us believe, and nationalisation would only leave staff, passengers and taxpayers worse off.
Metaphor	For	The current system has gone completely off the rails, but there is light at the end of the tunnel: we must go full steam ahead with nationalisation.
	Against	The franchise system has been chugging along nicely, and we should be clear that the likely consequence of nationalisation would be to create a complete train wreck.
Morality	For	The current franchising system, where corporations profit from taxpayer subsidies, is fundamentally wrong.
	Against	Many people never use the railways and so it is wrong to ask all taxpayers to pay for the costs of nationalisation.
Public Opinion	For	Poll after poll shows that people are dissatisfied with the train services they receive throughout the country.
	Against	The best sign of public support for the current franchise system is that since 1997 the total number of passengers on British railways has doubled.
Side Effects	For	One side-effect of nationalisation could be to eliminate regional differences in the quality of rail provision.
	Against	One side-effect of nationalisation of the railways would be to put all the financial risk onto the taxpayer.

Table A7: Treatment texts – Quotas for women on corporate boards

Element	For/Against	Text
Ad hominem	For	Those who oppose quotas are just afraid of letting women into their cosy boys' club.
	Against	Those who promote quotas are just virtue signalling.
Appeal to authority	For	A recent report by the European Institute for Gender Equality which endorsed quotas shows that there has been no progress in the UK for women in business over the past decade.
	Against	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".
Appeal to fairness	For	It is fair to increase the number of women on business boards so that everyone has an equal chance of getting these prestigious positions.
	Against	Like any other job, positions on business boards should be allocated fairly on the basis of qualifications, not quotas.
Appeal to history	For	The historical pace of change in women's representation at the top of businesses is too slow, we need to now take active measures to achieve equality.
	Against	History shows that quotas are unnecessary: there are now no all-male boards in the FTSE 100, compared with 21 such boards in 2011, and the percentage of women on FTSE 350 boards has more than doubled since 2010.
Appeal to national greatness	For	It should be a point of great pride that women in Britain are fundamental to our economic success, and adopting quotas would ensure that the UK is a great leader on this issue.
	Against	We should be proud of the extraordinary performance of the UK's many great businesses, and we should not risk that success with the imposition of unnecessary quotas.
	For	Quotas will not only help elite women, but also help ordinary women in normal jobs whose lives are shaped by the decisions of those on corporate boards.

Element	For/Against	Text
Appeal to populism	Against	We should be focussing on the position of ordinary working women, not on flashy quotas for the number of elite women on corporate boards.
Common sense	For	It is common sense that if businesses are failing to keep up with the modern world, you need to take stronger measures – such as implementing quotas – to help them catch up.
	Against	It is common sense that we should not interfere in who businesses put on their corporate boards.
Cost/benefit	For	Quotas have clear business benefits, as organisations with the highest level of gender diversity in their leadership teams are 15 per cent more likely to outperform their industry rivals.
	Against	The benefits of gender balance on boards are uncertain, as there is no definitive link between more gender diversity and stronger performance, but the costs are concrete, as business are more constrained in who they can hire.
Country comparison	For	We should follow the lead of the Scottish government, as they have committed to achieving gender balance on private, public, and third sector boards by 2020.
	Against	We should not make gender quotas an issue of law, but rather – like many other EU countries – allow businesses to take a voluntary approach to board and CEO recruitment.
Crisis	For	The current lack of women on boards represents a real crisis of wasted talent and potential.
	Against	Quotas are not necessary because there is no gender representation "crisis", and progress is being made and will be made over time.
Metaphor	For	Quotas will mean more powerful women on corporate boards who can act as beacons of light for other women in business.
	Against	Boardroom quotas are no more than a sticking plaster, and they will do nothing to address the root causes of women's underrepresentation in business.
Morality	For	It is a moral failure, in a society that aspires to be equal, to have such extraordinary low numbers of women in important positions.
	Against	Enforced quotas are immoral because they prevent hiring on the basis of individual merit.
Public Opinion	For	The UK public is strongly in favour of providing more opportunities for women to take on leadership roles in business.
	Against	In this country, the vast majority of women are not in favour of quotas.
Side Effects	For	One side effect of having more women on corporate boards is that there will be more role models for other women in business.
	Against	One unintended consequence of mandatory gender quotas is that they may create the perception that women on boards are not there by merit.

Table A8: Treatment texts – Reducing the legal restrictions on cannabis use

Element	For/Against	Text
Ad hominem	For	Those who oppose recreational cannabis use are mostly the type of people who never got invited to parties when they were young.
	Against	Many of those who campaign to legalise cannabis would really like to legalise all drugs.
	For	Transform, a charitable think-tank that campaigns for the regulation of drugs, argues that prohibition is a proven failure that will never be successful in protecting individuals or society from the misuse of drugs.

Table A8: Treatment texts – Reducing the legal restrictions on cannabis use *(continued)*

Element	For/Against	Text
Appeal to authority	Against	A report by the Royal College of Psychiatrists highlights that regular users of cannabis have double the risk of developing psychotic episodes or long-term schizophrenia.
Appeal to fairness	For	It is unfair to characterise responsible cannabis users as criminals.
	Against	Legalisation is unfair to the people who will be drawn into drug dependency.
Appeal to history	For	Looking back through the history of UK drugs policy reveals that criminalization does not work, we need fresh thinking and a new approach.
	Against	The terrible history of drug dependency in this country should be a stark reminder of the dangers of legalising drugs for recreational use.
Appeal to national greatness	For	The test of a great country is its ability to set sensible policies in difficult areas rather than trying to implement bans that cannot be effectively enforced.
	Against	Britain has led the world in introducing policies to reduce harmful drug use, we should not abandon these important values.
Appeal to populism	For	Elites already ignore drug laws, it is only ordinary people who ever get punished for using cannabis.
	Against	The elites who want to be able to smoke cannabis legally do not recognise the potential damage of more widespread use to ordinary people.
Common sense	For	It is common sense to legalise cannabis given that it is less dangerous than legal drugs like alcohol and tobacco.
	Against	It is common sense that we do not want people using cannabis and so it should remain illegal.
Cost/benefit	For	One of the biggest costs of cannabis prohibition is to enrich the organised criminals who provide drugs.
	Against	The productivity costs associated with widespread cannabis use would be substantial, there is good reason why it is not legal.
Country comparison	For	There is clear movement in many countries towards legalisation of cannabis: recreational use is now decriminalised in Canada, the Netherlands and also in some states in the USA.
	Against	There are very few countries in the world that have legalised cannabis.
Crisis	For	We should treat it as a crisis that we have criminalized large numbers of people for harmless use of cannabis.
	Against	There is no crisis to be solved by the legalisation of cannabis, but there might be one created by it.
Metaphor	For	People who use cannabis recreationally are not monsters; what is monstrous is a society that makes every small vice illegal.
	Against	Legalisation is a slippery slope, where we will start with cannabis and slide towards the gutter.
Morality	For	If it is morally permitted to get drunk in a pub, how can it be morally wrong to smoke cannabis in the privacy of your own home?
	Against	It is just wrong to use drugs to get stoned out of one's mind, it undermines one's self-control, leads to other wrongdoing, and above all it is morally degrading.
Public Opinion	For	According to a recent poll, the majority of the UK public back the decriminalisation of cannabis for recreational use.
	Against	Public opinion is firmly against the legalisation of cannabis for recreational use.
Side Effects	For	One attractive side-effect of legalising canabis is the additional tax revenue it would bring to the government.
	Against	An unintended consequence of legalizing cannabis would be to encourage children to try it.

Table A9: Treatment texts – Reducing university tuition fees

Element	For/Against	Text
Ad hominem	For	University Vice Chancellors are getting rich off the fees that young people pay and the debts that they take on.
	Against	The people who think you can slash fees without harming universities are just naïve.
Appeal to authority	For	The Institute for Fiscal Studies reports that 77 per cent of UK graduates will never pay off their full debt.
	Against	The current system works: the Institute for Fiscal Studies has shown that our universities are better funded than they have been at any point during the past 30 years.
Appeal to fairness	For	Current tuition fees are unfair for students from disadvantaged backgrounds, because they end up with the largest debts due to a lack of support for living costs.
	Against	Reducing tuition fees would be unfair to those who never got the advantage of a university degree, as they will have to pay extra tax to help university graduates improve their incomes.
Appeal to history	For	The current system means that young people have a worse deal than their parents and grandparents, who went to university for free.
	Against	We should not be nostalgic about the lower tuition fees of the past, which provided university education to many fewer students.
Appeal to national greatness	For	Great countries provide great education to all, they do not saddle young people with debt they will never repay.
	Against	The current tuition fee system has given us a terrific university sector in the UK that attracts huge numbers of fee paying international students.
Appeal to populism	For	High tuition fees cause no problems for the children of elites, but put ordinary young people deep in debt.
	Against	If we cut tuition fees, universities will need money from the public purse, so why should a taxi driver's taxes pay for a future banker's time at Oxford?
Common sense	For	The existing tuition fee system is not working as planned, and so it is simply common sense to make changes.
	Against	Common sense dictates that the people who get the benefits of higher education ought to be the ones to pay for it through their tuition fees.
Cost/benefit	For	The cost of budget tightening for universities pales in comparison to the benefits of reduced debt levels for students.
	Against	If the government cuts university fees without providing money to universities from tax revenues, there will be a devastating cost to the quality of education.
Country comparison	For	Students in England currently have the greatest amounts of student debt in the developed world – greater even than those in the United States of America.
	Against	Scotland has moved to a system of free higher education while maintaining bursaries for young people from disadvantaged backgrounds, allowing Scottish students to graduate with the lowest debt in the UK.
Crisis	For	The current tuition fee system is not even ten years old and we already have a crisis where many UK graduates are never expected to pay off their debts in full.
	Against	Cutting university fees without increasing taxpayer support would create a financing crisis in the higher education sector.
Metaphor	For	Young people are being crushed by the weight of debt created by current levels of university tuition fees.
	Against	Our universities are modern-day cathedrals, and reducing funding would erode the foundations on which they are built.
	For	It is wrong that under the existing system new graduates get a letter that effectively says: "Congratulations on graduating. Now we want the money back."

Element	For/Against	Text
Morality	Against	The current fees are based on the basic moral principle that students with degrees earn more money than those without, and so they should pay for their degrees.
Public Opinion	For	The National Audit Office reported that two thirds of students consider that universities do not provide good value for money.
	Against	Many people think that students should pay for their own education, and even more think that the taxpayer should not.
Side Effects	For	Declining mental health in students is an unintented consequence of the high levels of debt that students are taking on.
	Against	Since only high earners fully pay back their loans under the current system, lowering fees will have the unintended effect of primarily benefitting those who earn the most.

Table A10: Treatment texts – Renewing Trident

Element	For/Against	Text
Ad hominem	For	Only hippies and cowards think it is a good idea to unilaterally drop our nuclear deterrant.
	Against	The military industrial complex is lobbying hard to spend massive amounts of public money on a Trident replacement.
Appeal to authority	For	Experts within the UK armed forces emphasise the strategic importance of the Trident deterrent, particularly in our relationship with our NATO allies.
	Against	A recent report by Chatham House documents 13 incidents since 1962 in which nuclear weapons have nearly been used in error.
Appeal to fairness	For	Failing to renew Trident would be unfair to our children, who face an increasingly dangerous world.
	Against	It is not fair to spend vast amounts of money on our nuclear arsenal when that money is desparately needed elsewhere.
Appeal to history	For	Our Trident nuclear deterrent has been a vital part of our national security strategy for over half a century.
	Against	The Trident system was designed to deter the Soviet Union, which no longer exists.
Appeal to national greatness	For	The Trident system guarantees the UK a place at the world's top table.
	Against	Great Britain does not need status-symbol weapons such as Trident in order to hold on to our place as a leading nation in the world.
Appeal to populism	For	The only people who think that it is not important for UK to have a strong nuclear deterrent are metropolitan elites.
	Against	These kinds of weapon systems are just toys for elites, they are irrelevant to the concerns of ordinary people.
Common sense	For	It is common sense that we must have a strong nuclear deterrent in a world where nations like Russia and North Korea have nuclear weapons.
	Against	It is common sense to not spend a substantial fraction of our defense budget on weapons systems that will never be used.
Cost/benefit	For	Although it is expensive, the benefits of Trident are considerable: no alternative system is as capable, resilient or cost-effective as a Trident-based deterrent.
	Against	Replacing Trident would have considerable costs, as it would require an additional 6 per cent of the UK's defence budget, which will further reduce the amount we can spend on conventional armed forces.
Country comparison	For	We do not want to lag behind other major world powers – including Russia, the US, China and France – who are in the process of spending billions of pounds renewing their own submarine-based nuclear weapons.
	Against	Other countries – including South Africa, Brazil and Argentina – have made serious unilateral efforts to bring about nuclear disarmament, and we should join them.

A31

Element	For/Against	Text
Crisis	For	A submarine-based nuclear deterrant is the best defence against a potential nuclear crisis.
	Against	The crisis in our military is the decline of our conventional forces, and we should spend our money addressing this very real problem, rather than on renewing Trident.
Metaphor	For	Giving up our nuclear deterrent will leave us naked in global power politics.
	Against	Spending enormous amounts on Trident is like buying a tank to try to fight a swarm of mosquitoes.
Morality	For	If the consequence of possessing a nuclear weapon is that nobody else launches their own, and thus a conflict in which many millions would die is averted, then it is a moral imperative to possess that weapon.
	Against	Nuclear weapons are morally obscene: a technology that is capable of destruction and death at an indiscriminate and barbaric level.
Public Opinion	For	In poll after poll, two thirds of the British people endorse keeping and updating our Trident nuclear weapons system.
	Against	A recent survey suggests that the UK public are in favour of finding a cheaper way of keeping nuclear weapons, rather than renewing the Trident system.
Side Effects	For	One side effect of renewing the Trident system is that it will provide employment and economic benefits to many parts of the country.
	Against	One side effect of our continuing possession of nuclear weapons is to encourage other countries to maintain their own nuclear arsenals.

Table A11: Treatment texts – Spending 0.7 per cent of GDP on overseas aid

Element	For/Against	Text
Ad hominem	For	Opponents of overseas aid spending just do not care about the suffering of people around the world.
	Against	Supporters of overseas aid spending care more about people abroad than about people at home.
Appeal to authority	For	Oxfam says that overseas development aid has helped it to address several critical problems across the world.
	Against	The Overseas Development Institute stated that much of our aid money fails to promote peace and stability in poor countries.
Appeal to fairness	For	Overseas aid contributes to making the world a bit fairer for those who were unlucky to be born in poor countries.
	Against	Overseas aid is unfair to people in poor countries because it undermines the development of their countries by making them reliant on handouts.
Appeal to history	For	The UK has a long history of assisting countries that are desperately in need.
	Against	UK government spending on overseas aid is far higher now than at any point in the past and should be brought back down to historical levels.
Appeal to national greatness	For	The UK's commitment to overseas aid is part of what makes us who we are; it is part of the values of our country; it is part of what makes Britain great.
	Against	Great countries look after their own citizens: we should spend our money here rather than overseas.
Appeal to populism	For	We must show solidarity with the common people of other countries and do what we can to protect them from corrupt elites.
	Against	Up and down the country, ordinary people are asking "why are we spending our money on overseas aid, when children are going hungry here?"

Table A11: Treatment texts – Spending 0.7 per cent of GDP on overseas aid *(continued)*

Element	For/Against	Text
Common sense	For	If we want to help the neediest, it is common sense to do so in the poorest parts of the world, where even small amounts of money can make an enormous difference.
	Against	Common sense tells us that our first priority should be to help those in this country who are suffering.
Cost/benefit	For	Some of the benefits of UK foreign aid include providing nutrition to more than 28 million children and pregnant women, providing doctors for more than 5 million births, and providing 13 million people with emergency food assistance.
	Against	The 14 billion we spend on aid each year costs us the ability to invest at home on nurses, teachers and police.
Country comparison	For	The UK is one of six countries now meeting the UN target for overseas aid, which include Norway, Sweden and Denmark.
	Against	The UK spends far more than Japan, the United States, Italy, Portugal and Spain, each of which spend only 0.2 per cent of their GDP on foreign aid.
Crisis	For	Predictable aid flows allow agencies to put measures in place that mean that when crisis or disaster strikes, the resources are there to be mobilised immediately.
	Against	Our own social safety net is currently in crisis, we can ill-afford to send more money abroad.
Metaphor	For	The money we spend on foreign aid is the lifeblood of development in countries across the world.
	Against	Money spent on foreign aid may just as well be poured down the drain.
Morality	For	Spending on overseas aid allows us to fulfill our moral duty to support those in need around the world.
	Against	In our effort to spend enough aid money to meet a fixed target, we end up supporting immoral causes.
Public Opinion	For	Repeated opinion polls show that a majority of people in the United Kingdom support spending on overseas aid.
	Against	More than 100,000 people signed a petition to call on the government to reduce spending on overseas aid.
Side Effects	For	One positive side-effect of our investment in overseas aid is that it helps to stengthen our diplomatic ties with many countries around the world.
	Against	Overseas aid often has unintended consequences, for example if you deliver free food to a country, it makes it difficult for farmers in that country to make a living.

Table A12: Treatment texts – Sugar tax in the UK

Element	For/Against	Text
Ad hominem	For	Large confectionary companies are spending millions to make sure the government does not adopt a sugar tax.
	Against	Sugar tax supporters think that other people are too stupid to make healthy choices themselves.
Appeal to authority	For	Public Health England suggests that a price increase on high sugar products would lead to a decrease in sugar consumption.
	Against	Public Health England have concluded that a sugar tax on its own will have a limited effect in reducing the nation's sugar intake.
Appeal to fairness	For	A sugar tax is important because it helps to all children a fair chance of a healthy life.
	Against	It would be fairer if we subsidised access to good food rather than putting up taxes on bad food like sugar.
	For	There is plenty of historical evidence that Government action can change people's behaviour, and a sugar tax would be no different.

Element	For/Against	Text
Appeal to history	Against	The experience of recent years provides an argument against a sugar tax, as voluntary action by industry has resulted in tonnes of sugar being removed from products without any need for new taxes.
Appeal to national greatness	For	If a sugar tax is implemented, the UK would become a world leading role model for other countries seeking to improve public health.
	Against	The great people of our country should be trusted to make their own decisions – they do not need a tax on sugar to tell them what is healthy.
Appeal to populism	For	A sugar tax makes business elites pay for the costs of childhood obesity that they have done so much to encourage.
	Against	Taxes on sugar will inevitably hit ordinary people harder than avocado-eating metropolitan elites.
Common sense	For	It is just common sense that if a sugar tax is introduced, people will consume less sugar and the government will both gain tax revenue and save money on the health service.
	Against	Common sense dictates that the state should not use taxes as a way of telling people how to live their lives.
Cost/benefit	For	A 20 per cent sugar tax would raise about 1 billion, which would benefit areas of the budget that are currently being cut.
	Against	A 20 per cent sugar tax would raise about 1 billion, a cost that would fall disproportionately on the poor, who spend a larger share of their money on food and drink.
Country comparison	For	There is evidence from Mexico and France that when a sugar tax is implemented, people's behaviour starts to change and they start to choose sugar-free alternatives.
	Against	There simply is no good evidence from other countries that sugar taxes have the desired effect on health.
Crisis	For	A sugar tax could help us to address the current obesity crisis, which is marked by the fact that nearly half of children are now overweight.
	Against	A sugar tax would be yet another example of the increasing interference of the government in everyday life, which is a crisis that we must address.
Metaphor	For	Big sugar has its hand on the throat of the Government and it is big sugar that determines policy.
	Against	A sugar tax is yet another example of the nanny state telling us what to do.
Morality	For	For the government to fail to address the dangers of high sugar consumption would be a great moral dereliction of duty.
	Against	It is simply wrong for the government to use a sugar tax to interfere with what people choose to eat and drink.
Public Opinion	For	Recent opinion polls suggest that a majority of the public are in favour of a sugar tax.
	Against	Recent opinion polls suggest that a majority of the public are against a sugar tax.
Side Effects	For	One potential side-effect of a sugar tax would be to reduce the amount that the government has to spend each year on obesity treatments.
	Against	A likely unintended consequence of a sugar tax would be for companies to introduce other unhealthy sweetening agents into our drinks.

References

Coppock, Alexander, Thomas J Leeper and Kevin J Mullinix. 2018. "Generalizability of heterogeneous treatment effect estimates across samples." *Proceedings of the National Academy of*

Sciences 115(49):12441–12446.