Measuring the Persuasive Power of Political Arguments *

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

Despite the central importance of persuasion in modern politics, there are few existing studies that characterise the relative persuasiveness of different kinds of political arguments. One way that political arguments differ is in the rhetorical elements that they use: arguments may be based on appeals to relevant authorities, comparisons to other countries, they may use metaphors or be ad hominem. We use a novel strategy for assessing the relative efficacy of these rhetorical elements without reliance on a single implementation by using a large number of treatment implementations that are individually underpowered but are collectively informative about the average and distribution of effectiveness. We find that there are modest differences in the average effectiveness of different rhetorical elements and large differences across different arguments using the same element type. Our approach also allows us to assess argument strength along other dimensions: we find that arguments that are relatively persuasive when evaluated against arguments on the same side of an issue are also relatively persuasive against arguments on the other side as well finding that individuals with substantially different political views still evaluate argument strength in broadly similar ways. Our results have important implications for the interpretation of previous experimental studies as well as the design of new studies in the field of political communication.

Introduction

How persuasive are different types of political argument? Politicians invest time and effort in crafting arguments to present to voters, and the arguments they make across multiple issues often share similar rhetorical elements. Regardless of the specific policy at stake, politicians can draw on endorsements from relevant authorities; emphasise the moral rationale for supporting the policy; carefully articulate costs and benefits; launch attacks that impugn the motives of opposition actors; present evidence from historical experience; and so on. While interest in the the rhetorical strategies that political figures use has sustained over the course of millennia (Aristotle, c.322 BCE; Rhetorica ad herennium, c.80 BCE; Riker, 1990; Charteris-Black, 2011), and

*This version: August 30, 2019

more recent work has begun to test the efficacy of different communication strategies (Thibodeau and Boroditsky, 2011; Lau, Sigelman and Rovner, 2007; Bos, Van Der Brug and De Vreese, 2013; Atkins and Finlayson, 2013; Jung, Forthcoming; Jerit, 2009; Arceneaux, 2012), empirical study in this area remains limited. In particular, empirical research has only been able to make very narrow comparisons of different types of arguments that politicians employ in particular domains. A consequence of this is that "scholars still understand little about the factors that shape argument strength" Arceneaux (2012, 272).

Determining whether some (and which) types of argument are more successful than others is an important endeavour from a normative perspective. 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 (Madison, 1788; Kant, 1790; Elster, 1998). More recent work in normative political theory which attempts to "rehabilitate rhetoric" (Chambers, 2009; Dryzek, 2010; O'neill, 2002) also suggests that, while rhetoric may not be damaging *per se*, specific forms of rhetoric – particularly when used communicate "vapid and vacuous" statements rather than substantive policy information – should still be viewed as "a threat to deliberative ideals." (Chambers, 2009) For instance, when voters 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, when voters are persuaded by arguments that rely on appeals to factual information and expert authority, we might be more sanguine about the effects of political communication.

In this paper, we provide the first quantitative evaluation of the relative effectiveness of different rhetorical elements by introducing a new experimental measurement strategy and associated modelling approach. For the purposes of our experiment, we take a broad view of what constitutes political rhetoric. Following canonical work in the literature, we think of rhetoric as a "range of methods for persuading others" (Charteris-Black, 2011, 13) that includes several commonly employed elements that we outline below.[1] Our enumeration of rhetorical elements is not exhaustive, and does not correspond directly to existing rhetorical typologies in the literature

[1]We share the understanding of Atkins and Finlayson (2013, 161) who suggest that analyses of political rhetoric should focus "on the varied kinds of proof or justification found in political argument." Likewise, our decision to focus on styles that arise regularly in UK politics is consistent with calls to study "political arguments as they take place 'in the wild' " (Finlayson, 2007, 552).

(e.g. Aristotle, c.322 BCE; Charteris-Black, 2011; Finlayson, 2007). Instead, our design is informed by the types of argument made in contemporary British politics, and especially in speeches delivered by MPs in the UK House of Commons.[2] Our 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. The core of our experiment consists of presenting pairs of these arguments to survey respondents and asking them to assess which of the pair is most persuasive. We then use the distribution of responses to these forced-choice comparisons to generate estimates of the persuasive power of each of the arguments and, in turn, of the average persuasive power of each of the rhetorical elements.

There are three advantages of our experimental design which are useful in our application and which would be useful in a range of other survey experiments as well. First, by specifing multiple implementations of each of our rhetorical elements, we reduce the likelihood that our results will be confounded by specific features of the treatments that implement those elements. In a recent paper, Grimmer and Fong (2019) illustrate that an endemic problem in vignette experiments employing single-text treatments is that researchers rely on an implicit and implausible exclusion restriction: that the latent treatment of interest (here, rhetorical element) is uncorrelated with other implementation-specific features. Our approach – which includes 24 separate treatment texts for each of our 14 rhetorical styles – significantly reduces the likely magnitude of such confounding and enables post-experimental checks related to specific confounding concerns. Second, in contrast to the vast majority of existing literature, we test for rhetorical effectiveness across many policy domains at once. This increases the external validity of our estimates relative to those in the existing literature as it allows us to describe not only the average effect of each rhetorical element, but also the variance of those across implementations and policy issues. Finally, we allow for comparisons of effectiveness between pairs of arguments when both arguments are on the same side of an issue as well as between arguments on different sides of an issue. This design choice allows us to assess concerns articulated by Graham and Coppock (2019) that survey respondents are likely to overstate changes when they are asked to give self-assessments of the effects of an experimental treatment on their political attitudes.

[2] As we discuss in the next section, the styles we identify in many cases relate to ongoing debates in a diverse contemporary literature focussing on political communication.

The structure of the data that comes from our experiment suggests the use of a measurement model for paired comparisons. A natural choice for this task is a Bradley-Terry model (Bradley and Terry, 1952), which can be used to produce estimates of the relative persuasiveness of each of the arguments we include in the experiment. However, unlike previous use of such models to assess argument strength (Loewen, Rubenson and Spirling, 2012), our main quantity of interest is not the efficacy of individual *arguments*, but rather the distribution of effectiveness of many individual arguments deploying each type of *rhetorical element*. We therefore incorporate a Bradley-Terry style model for the experimental response data within a series of hierarchical models that describe the variation in argument effectiveness. This approach enables us to make statements about differences in average effectiveness of element types as well as assessing effect heterogeneity among arguments deploying the same element. We explicitly model the effect of arguments being on the "for" or "against" side of an issue, which allows us to distinguish between arguments' relative persuasive power and the degree to which respondents tend to agree with the side of the issue on which those arguments appear.

The modelling framework we propose can be adapted to address other questions of interest. For example, we show that using this setup we are able to (1) control for other features of our treatment texts that might potentially confound our causal effects of interest; (2) assess the possibility that arguments have different relative strengths when facing other arguments on the same side of the issue than when facing argument on the opposing side of the issue; and (3) assess the degree to which different groups of respondents perceive the relative persuasiveness of different arguments differently. Although we use this framework for analysing rhetorical elements in the context of UK politics, both our experimental design and associated modelling strategy could easily be applied in a wide range of applications across the study of public opinion, political behaviour and political communications.

The main findings of our study are as follows. First, we find that there are modest average differences between arguments with different rhetorical elements. However, those differences that do emerge clearly from the data are related to current concerns about the quality of democratic debate and political communication in the UK. We find that the strongest rhetorical element in our experiment are *appeals to authorty* – 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 and authority in political debate became a prominent issue in UK politics[3] during the Brexit referendum in 2016 when Michael Gove, a leading figure in the "Leave" campaign, declared that the public "have had enough of experts".[4] Recent descriptive work has also documented a decline in authority-based appeals of this type in UK politics (Atkins and Finlayson, 2013). Our results suggest that, despite this decline, making appeals to relevant figures of authority remains a persuasive way 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.

Second, and in some sense more importantly, we find that the heterogeneity in the effectiveness of different implementations of these rhetorical elements is much larger than the average differences. For example, we find that while *appeals to authority* are on average more persuasive than other rhetorical styles, some appeals of this sort are still amongst 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. 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 (Bos, Van Der Brug and De Vreese, 2013, Arceneaux (2012), Jung (Forthcoming), Nelson (2004), Jerit (2009)). Our results imply that, in addition to traditional concerns about the representativeness of survey respondents, an additional external validity concern for studies such as these is that the treatment effects that they estimate may

[3] See, for example, Should we listen to the experts on the EU referendum, The Guardian, 8 June 2016; Michael Gove was (accidentally) right about experts, Spectator, 14 January 2017; and Have we fallen out of love with experts?, BBC, 27 February, 2017

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

not generalise to other policy domains, or even to other implementations of the same concept in the same domain.

Finally, we demonstrate two findings regarding the relative strength of the arguments that are not focused on the rhetorical elements. First, we show that the estimated relative strength of arguments based on respondents evaluating persuasiveness of two arguments on the same-side of an issue are highly positively correlated with those we estimate based on opposite-side comparisons. Same-side comparisons also yield more variation in the evaluated strength of arguments, which combined with the high, but imperfect, correlation with opposite side evaluations means that they provide a similarly strong signal regarding opposite side comparisons as do the opposite side comparisons themselves. Second, we show that there is a large degree of correlation in the effectiveness of political arguments across different types of respondents. While Labour and Conservative, Leave and Remain voters evaluate arguments somewhat differently, the effectiveness of different arguments for these groups are still broadly positively correlated with one another across the range of arguments we implemented. In contrast to the high degree of heterogeneity in the strength of different arguments, the degree of heterogeneity in individual responses to those arguments is modest.

Rhetoric, persuasion, and public opinion

Which types of political rhetoric are most effective at shifting public opinion? In sections below we outline 14 rhetorical "elements" that we identify as features of contemporary UK political debate. As we describe below, our goal in defining these elements is to capture a wide range of behaviours that occur frequently in UK politics, and our primary interest is in assessing the relative persuasiveness of these elements. However, we note here that many of the elements that we identify relate directly to ongoing literatures in political behaviour and political communication.

For instance, one rhetorical element that we evaluate is the use of *metaphor*. 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), though recent experimental evidence suggests that the causal effects of analogous reasoning on public attitudes are more limited (Barnes and Hicks, 2019).

Similarly, we also include treatments which relate to the idea of *ad hominem* attacks. There is an extensive literature which examines whether negative political "attacks" can be persuasive communication strategies, and summaries of the available empirical evidence suggests that, in general, these types of attacks are not effective in changing public opinion towards specific candidates or policies (Lau, Sigelman and Rovner, 2007).

We also engage with debates over the effectiveness of *populist* rhetoric, which has been a source of increasing study in recent years. Descriptive analysis suggests that there has been a "populist shift" in the rhetoric of UK political actors in recent years (Atkins and Finlayson, 2013), and studies suggest that rhetoric that casts political conflict as a contest between the "people" and the "elite" can influence perceptions of the legitimacy of political actors (Bos, Van Der Brug and De Vreese, 2013) as well as attribution of blame to different levels of government (Hameleers, Bos and de Vreese, 2017) and demographic groups (Hameleers and Schmuck, 2017). We include "populist" language as one of our styles of interest.

Jung (Forthcoming) suggests that the use of *moral* rhetoric, which focusses on the degree to which political actors characterise their positions as being justified by moral beliefs about right and wrong (as opposed to, say, more pragmatic policy concerns), activates emotional responses from respondents and makes voters more likely to mobilize during elections. Relatedly, Nelson (2004) suggests that by strategically prioritizing certain moral values, politicians can influence public opinion on policy issues. In addition to testing the efficacy of moral arguments invoking "right" and "wrong", we have a separate category that invokes arguments on the basis of "fairness".

There is a lengthy literature examining the role of "cues" in political communication. A cue, broadly defined, is a piece of information that allows voters to make inferences about the desirability of a policy or a candidate without having to develop detailed knowledge. Cues could include endorsements by prominent figures; information about the level of public consensus on an issue; or information on partisan stances. Endorsement-based messages, for example, have

been shown to increase support for referenda initiatives, a finding that holds both for partisan (Boudreau and MacKenzie, 2014) and non-partisan (Dewan, Humphreys and Rubenson, 2014) cues. There is also qualitative evidence to suggest that appeals to authority and expertise have become less frequent in recent years in UK political discourse (Atkins and Finlayson, 2013). We include two distinct categories of cues in our experiment by testing whether "appeals to authority" and "appeals to public opinion" are considered persuasive.

Another rhetorical tactic employed by politicians is to strategically communicate the expected *costs and benefits* of a particular policy. These types of "predictive appeals" (Jerit, 2009) can be used to generate support for a policy by suggesting that there are overwelming gains to be accrued by adoption, or unsustainable losses that would arise if the policy is not adopted. As William Riker argued, "campaigners on each side emphasize the dreadful consequences of the failure (or success) of the motion they advocate (or oppose)." (Riker, 1990, 58) We incorporate these considerations into our experiment through rhetorical elements that invoke "cost vs benefit arguments" and separately through those that identify "side effects" of different policy options.

We see our approach as offering three primary contributions to these diverse literatures on different rhetorical elements. First, in almost all of the papers cited above, the persuasiveness of the relevant style of interest – populism; metaphor; morality; etc – is evaluated in the context of vingette experiments where a treatment text containing the relevant element is contrasted with a control condition that does not include that element.[5] Importantly, very few of these papers *compare* the element of interest with other plausibly applicable rhetorical elements. As a consequence, these papers give information about the use of each element in isolation, and often versus no information at all. This is unfortunate, as it makes it difficult to make statements about the relative effectiveness of different elements. As one recent review of the empirical literature on persuasive effects argues, "the estimates of the effectiveness of persuasion…are hard to compare." (DellaVigna and Gentzkow, 2010, 645)

From a normative perspective, it seems important to determine not only whether individual rhetorical elements can be effective versus no argument at all, but also which are more effective

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

relative to one another. For example, knowing that populist rhetoric can affect views about the legitimacy of different policy alternatives is possibly less troubling if these types of effects are swamped by larger effects of more substantive rhetorical elements such as a policy's specific costs and benefits. However, existing work in this literature largely avoids combining different rhetorical elements in the same study, which makes comparisons of the persuasive efficacy of different elements impossible. Our appraoch, by contrast, allows us to directly assess the relative persuasive power of different rhetorical elements.

Second, the literatures outlined above generally present evidence of rhetorical effects on the basis of experiments that pertain to very few policy issues. A natural concern is therefore whether the effects detected in these experiments generalise to the broader (albeit difficult to specify) population of political issues for which politicians might use these types of rhetoric.[6] 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 to immigration? Are metaphors as important in political arguments about policy areas other than crime (Thibodeau and Boroditsky, 2011)? 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)? Single-context studies are not informative in this regard.

The approach we suggest helps to address these problems, and allows us to characterise the degree of heterogeneity in persuasive effects of these types of rhetoric across issues. In our set-up, each latent concept of interest – populism; metaphor; cost/benefit; morality; etc – is represented by many different treatment texts across many different policy areas of potential interest. Although each of these treatments is individually underpowered, they can be used

[6]The desire of scholars to generalise in this way is evident from the concluding claims authors provide. For instance, in their study on the effects of metaphor in political communication, Thibodeau and Boroditsky (2011) construct treatments that relate only to the issue of crime but conclude that "far from being rhetorical flourishes, metaphors have profound influences on how we conceptualise and act with respect to important societal issues" (Thibodeau and Boroditsky, 2011, 1). Similarly, Bos, Van Der Brug and De Vreese (2013, 204), using an experiment which focusses on the narrow issue of whether the government should build new nuclear power station, summarise their results of their experiment as showing "that using a populist style does positively affect the perceived legitimacy of right-wing populist party leaders". Likewise, Arceneaux (2012, 281) makes the general point that "arguments that claim to avert losses have the potential to be more persuasive than those that claim to realize gains" on the basis of one experiment that relates to a fictional policy issue surrounding an Asian bird flu outbreak and a second experiment related to an issue of freedom of speech in high schools.

collectively to make inferences about the average persuasive power of arguments of that type, and also the degree of treatment effect heterogeneity across issues. In contrast, no matter how large the sample is for a traditional experimental design for a single implementation, it can never tell us anything about whether that implementation is representative of the more general phenomenon of interest. Indeed our results suggest that we should not place undue weight on general claims drawn from individual implementation experiments: we uncover significant treatment effect heterogeneity of arguments that share a common type.

Third, our design helps to alleviate an additional problem that is common to vignette-style experiments that use single-text treatments as the basis for inference. Grimmer and Fong (2019) show that random assignment of texts to respondents only helps researchers to overcome the problem of the selection of units into treatments, but does not help with the fact that "latent treatments within the texts may co-occur in patterns that confound inference." (p.3) If the latent treatment of interest to the researcher correlates with other textual features then treatment effects based on these texts cannot necessarily be attributed to the effects of the latent treament, as they might reflect instead the effects of the other correlated features.

Grimmer and Fong (2019, 20) suggest that one solution to this problem is to provide several texts per latent treatment, "because it enables scholars to vary the background features" which might confound the latent concept of interest. If background features vary independently across texts in which the concept of interest remains constant, then researchers can average over the effects of these separate treatments and attribute the average effect to the latent concept. That is, in the same way that randomization of units to treatments allows researchers to average over potentially confounding unit-specific characteristics, providing multiple treatments per concept of interest allows researchers to average over potentially confounding treatment-specific characteristics.[7]

An example helps to clarify the intuition. Imagine we design an experiment to detect the effects of "emotional" appeals on the level of support for a given policy. We therefore write two treatment texts that are designed to include emotional language. In writing these two texts, we inadvertently use "positive" emotional language in the first (expressing some positive conse-

[7] Of course, the crucial assumption here is that all potentially confounding background features of texts do indeed vary *independently* across the different treatment texts.

quence of adopting the policy) and "negative" emotional language in the second (experssing a negative consequence of failing to adopt the policy). Were we to use only one of these treatments as the basis of our experiment, it is clear that our latent feature of interest ("emotion") would be confounded by the addition of an unintended latent feature in the text ("tone"). However, if we include both texts in our experiment, we can marginalise over the sentiment conditions to recover the effects for emotion, our concept of interest.

While we build on the intuition and formalization presented in Grimmer and Fong (2019), our study differs from theirs in terms of research design and empirical application. In their paper, the empirical analysis focusses on estimating the casual effects of latent treatments that are discovered via a machine learning technique (Fong and Grimmer, 2016) applied to a corpus of tweets sent by President Trump. Our approach is relevant to more common forms of experimental research where scholars typically have a fixed (though latent) concept that they wish to investigate, and they construct treatments that implement that concept. Below, we describe the 14 rhetorical elements that are our primary latent treatments of interest, and the 24 individual treatment texts we use in the experiment for each element

In addition to our design being a closer approximation to the modal social science experiment of this type, our results highlight an important point about the heterogeneity of treatment effects of texts that at least ostensibly represent the same latent concept. As we discuss below, we find that there is often substantially more variation in treatment effects within rhetorical types than between them.

From a modelling perspective, the paper closest in spirit to our work is by Loewen, Rubenson and Spirling (2012), who evaluate the relative persuasive power of 12 arguments relating to an electoral reform referendum in Ontario, Canada, in 2007. As with their work, we ask survey respondents to make forced-choice comparisons between pairs of political arguments and use the responses to estimate of the persuasive power of those arguments via extensions to a Bradley-Terry model. However, our project differs from theirs in several ways.

First, our primary quantity of interest is not in the relative persuasive power of individual arguments, but rather the power of different rhetorical element types which can appear in many different arguments. For example, we are interested in whether arguments that make appeals

based on "crisis" considerations are on average more persuasive than arguments that make appeals based on "morality" considerations, rather than whether any specific "crisis" and "morality" arguments that relate to a particular issue are more persuasive. One consequence of this is that we base our study on a much larger number of arguments – we include 336 individual treatments across 12 distinct issue areas – which while individually underpowered are nonetheless informative about the relative persuasive power of the different rhetorical styles.

Second, a limitation of the approach outlined by Loewen, Rubenson and Spirling (2012) is that it cannot be used to distinguish between the persuasive power of an argument and the tendency for survey respondents to approve of the side of an issue on which the argument appears. For example, their results (p. 218) show that "all else equal…an argument in favour of the status quo appears systematically advantaged". Although politicians clearly have incentives to engage in strategic position taking in order to attract support, once a policy position has been adopted, the key task of political orators is to make the most persuasive argument possible for that position regardless of which side of the issue it is on. A helpful feature of our measurement strategy is that is allows us to distinguish the relative persuasiveness of arguments from the relative popularity of the issue positions that they endorse.

Experimental design

In our experiment (fielded by YouGov to their UK online panel in June 2019), we presented respondents with pairs of arguments pertaining to current issues in UK politics and asked them to select, for each comparison, which of the two arguments was more persuasive. Before discussing how we approach analysing this data, 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, ranging from "Nationalisation of the railways in the UK", to "Reducing university tuition fees", to "Spending 0.7% of GDP on overseas aid". The full list of policies we include in our design can be found in table 2 in the appendix, along with the description of each issue that we

use in the experiment. These represent a selection from the broader set of issues discussed in UK politics in recent years. In deciding which policies to include, we focused on identifying those where there were clear political divisions, both amongst politicians and amongst the public, but where these divisions were not too strong.

A **rhetorical element** is a feature of political argument that is used to emphasise the desirability or undesirability of a given policy. Our experiment investigates the efficacy of 14 rhetorical elements, including those that rest on appeals to various forms of evidence (such as "Cost vs benefit arguments", "Country comparison", and "Appeals to history"), elements that are more directed at communicating values (such as "Appeals to fairness", "Morality", and "Appeals to national greatness"), and more classical forms of rhetoric (such as "Metaphor/figures of speech", "Ad hominem", and "Crisis"). The full set of rhetorical elements that we consider, as well as descriptions of the content of each type, are given in table 1.

Our goal is to assess the effectiveness of the types of arguments made by politicians in modern politics, and we therefore based our categorisation of rhetorical elements on close reading of contemporary political debates, rather than on existing accounts of political rhetoric. We began with a short list of possible rhetorical categories, and then we expanded and refined our categorisation by searching through transcripts of debates in the UK House of Commons that related to the issues defined above. An alternative approach would have been to base our design on a pre-existing decomposition of either contemporary (e.g. Charteris-Black (2011)) or classical (e.g. Aristotle (c.322 BCE)) rhetorical appeals. However, our primary interest is in quantifying the persuasiveness of the types of rhetorical appeals that we actually observe being used in contemporary politics, rather than those promoted in classical antiquity.

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.

An **argument** is a text that makes a case in favour or against a specific policy. While arguments could include any number of rhetorical elements, for purposes of our experiment we exclusively designed arguments that used a single element from the typology we developed. For each policy issue in our experiment, we wrote two separate arguments for each of the rhetorical elements: one arguing in favour of the policy, and one arguing against the policy. For example, for the "Reducing the legal restrictions on cannabis use" policy issue and the "Crisis" rhetorical style, we have:

- 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."

where the first argument employs the idea of crisis to encourage support for legalisation, and the second uses the idea of a crisis to encourage opposition to legalisation.

Similarly, for the "Sugar tax in the UK" policy and the "Morality" style, we have:

- 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."

Here, the first argument suggests the moral imperative is to take action on sugar consumption, and the second identifies a moral imperative to avoid government interference on this issue.

We construct arguments of this type, for each rhetorical style, both for and against each policy issue. This results in 2 x 14 x 12 = 336 separate arguments. These arguments form the basic treatments in our experiment. To ensure that the arguments we use in the experiment resemble the types of argument used by politicians in UK politics, we searched through the transcripts of debates held in the UK House of Commons that pertained to the policy issues outlined above. From these debates, we extracted sentences and paragraphs that corresponded to our rhetorical styles, and then subsequently edited these texts into the form we use in the experiment. Tables

3-14 in the appendix show all 336 arguments we include, and for many of the sentences we provide hyperlinks to the source documents on which our treatments are based. In some cases it was not possible to identify an example of our rhetorical styles in the texts of the Commons' debates on a particular issue, and so for these cases we wrote arguments of our own, ensuring that the texts were as similar as possible in style to those based on politicians' speeches.

While this procedure ensured that the arguments we presented are all arguments that UK politicians could plausibly make, the arguments we developed do not constitute a random sample from a well-defined population of possible arguments. We cannot claim that our implementations of each element type are representative of that type more broadly either. These are limitations that are just as present in single-implementation studies as in our design, they are just less obvious. As we discuss later, having variation in implementations, even if it is not representative of a target population distribution, enables valuable analyses that are not otherwise possible.

Survey instrument

We use these arguments as the basis of a forced-choice experiment, in which respondents are randomly presented with two arguments pertaining to a particular policy issue and asked which of the two arguments they find the most persuasive. Respondents to our survey were first presented with an introduction screen which outlined the basic idea behind the experiment. They were told that they would be asked about "some issues that feature in UK politics", that for each issue they would be presented with two arguments, and that they would be asked to select which argument they find most persuasive.

Each respondent was asked to make four argument comparisons, each focussed on a different policy issue. The four issues for each respondent were randomly sampled with equal probability without replacement from the total population of 12 policy issues. For each selected policy, we then randomly sampled whether a respondent was presented with two arguments "in favour" of that policy, two arguments "against" that policy, or one argument "in favour" and one argument "against". In 50% of cases, respondents were presented with two arguments on the same side of a policy, and in 50% of cases they were presented with arguments from alternative sides of the

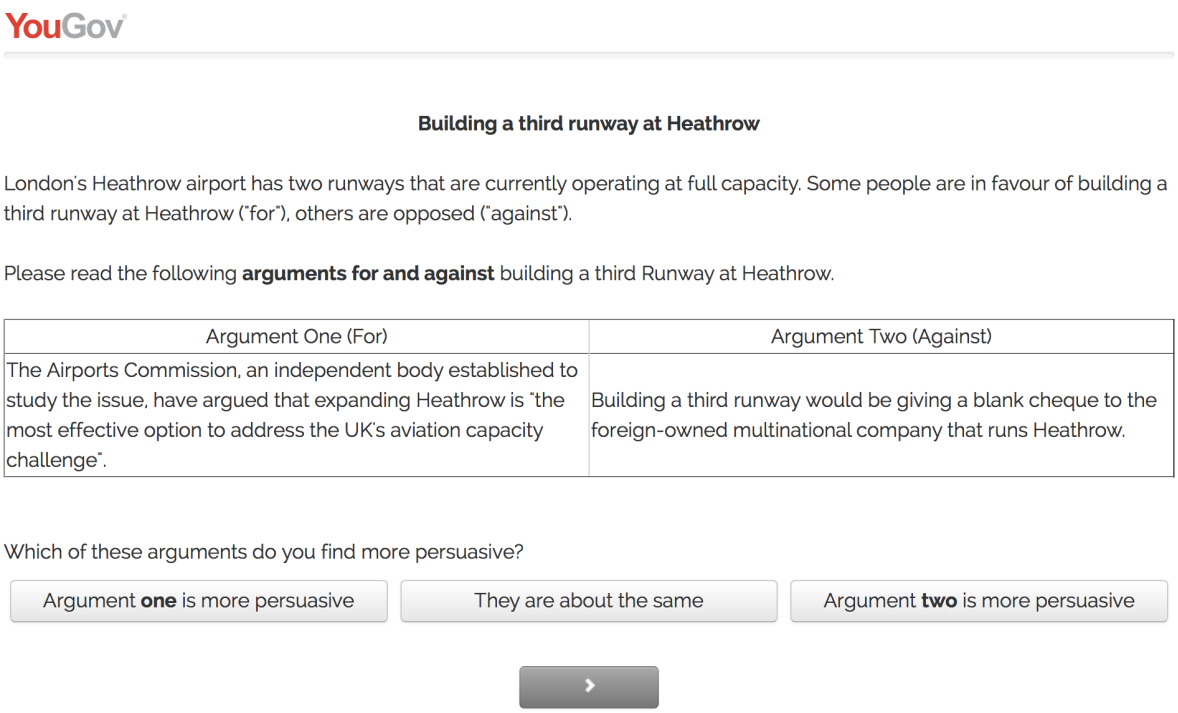


Figure 1: Experiment prompt

policy.

The main question screen of the survey instrument used in the experiment is given in figure 1. We provide respondents with the policy issue title (in the example here, "Building a third runway at Heathrow"), a short description of the policy and the context of the political debate, and an instruction to read the [two arguments in favour/two arguments against/arguments for and against] the policy issue in question. The two arguments for each comparison were presented to respondents side-by-side on the screen, and the arguments were labelled as being either for or against the policy issue.

As the wording of the survey prompt clearly reflects, we designed our study to assess "persuasiveness" rather than "persuasion". That is to say, we wanted to look at self-reported assessments of arguments by respondents rather than the treatment effects of different arguments on respondents' own positions. While there are many applications where persuasion rather than persuasiveness is the quantity of interest, focusing on the latter allows us to generate a novel comparison of the performance of arguments in comparisons to other arguments on the same side of an issue versus those on the opposite side of the issue. In opposite side comparisons

we might expect that some respondents will simply select the arguments which correspond to their own prior preferences about a given policy issue. For example, a respondent who strongly believes that the railways ought to be renationalised may consider any single sentence arguments against nationalisation less persuasive than any single sentence argument in favour. By contrast, the same respondent presented with two arguments on the "against" side may be able to provide useful information about which of those alternative rhetorical strategies is more persuasive.[8] An advantage of including both types of comparisons is that we are able to use the results to also evaluate the degree to which "same side" and "different side" evaluations result in similar rankings of arguments. This is something we describe in more detail below.

We collected 4 responses each from 3317 respondents, giving us a total of 13268 observations. Given the (336) individual treatments, each treatment appears in an average of 79 pairwise comparisons in our data (sd = 8.6). Thus this sample size does not give us sufficient power to estimate precise effects at the level of individual arguments. However, our main quantities of interest are not the effects of individual arguments, but rather the average effects of each of our rhetorical elements as well as the degree of heterogeneity in across different arguments deploying the same type of element, and we do have sufficient data to learn about this distribution.

Modelling Persuasive Arguments

Our design generates a set of responses that specify "winners" from a pairwise competition between two arguments, with the possibility of ties. Overall, we have a set of 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 that come from our experiment. We assume that our arguments can be described in terms of their latent and unobservable persuasive "power", where the power of a given argument is denoted α_j. The Bradley-Terry model for paired comparisons (Bradley and Terry, 1952) models the log-odds that argument j beats argument j' in a paiwise comparison as:

[8] For some applications we might not care about which arguments such a person finds more persuasive, because they will never be persuaded.

$$log\left[\frac{p(j \text{ wins})}{p(j' \text{ wins})}\right] = \alpha_j - \alpha_{j'}$$

Each argument j is described by a single "strength" parameter α_j. The stronger argument j is relative to argument j', the higher the probability that argument j beats argument j' in pairwise comparison (Loewen, Rubenson and Spirling, 2012). The Bradley-Terry model does not allow for the possibility of ties, which are present in our data:

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

Several extensions of the Bradley-Terry model that account for the possibility of tied outcomes in comparisoins exist (Rao and Kupper, 1967; Davidson, 1970; Glenn and David, 1960; Agresti, 1992; Tutz, 1986) which generalise the Bradley-Terry model to outcomes with 3 (or more) ordered response categories. We adopt a model of the following form:

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

where θ_k is the cutpoint for response category k and α_j is an argument random effect which represents the persuasive power of argument j.[9] If we had only a few arguments to test and a very large number of responses involving each one, the α_j would be our primary quantity of interest and we could simply use this as our full model specification. Instead, we are interested in modelling how the strength of arguments α_j vary as a function of features of those arguments. With the core specification of equation 2 in hand, we can specify various different hierarchical models for the α_j parameters that allow us to address different research questions.

[9] If $\theta_1 = \theta_2$, the probability of the "About the same" response goes to zero, and we recover the Bradley-Terry model.

The Relative Persuasiveness of Rhetorical Elements

Our primary quantity of interest in this study is the distribution of the strength of arguments using each of the 14 rhetorical elements. 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:

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

$$\mu_e \sim N(0, \omega) \tag{3}$$

$$\nu_j \sim N(0, \sigma_{e(j)}) \tag{4}$$

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 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 the distribution of argument-level effects around the element-type average, we estimate separate variance parameters for each element ($\sigma_{e(j)}$).[10]

Figure 2 shows the estimated average strength μ_e for each of our 14 rhetorical elements as well as the estimated relative persuasive strength $\mu_{e(j)} + \nu_j$ of each of the 336 individual arguments that we include in the experiment. Blue numbers 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. For instance, the figure shows that the strongest argument in our experiment is the "Appeal to au-

[10] 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.

thority/endorsement" argument which is against the introduction of quotas for women on corporate boards.[11] By contrast, our estimates suggest that the weakest of our arguments is the ad hominem attack in favour of renewing Trident.[12]

[11] "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'."

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

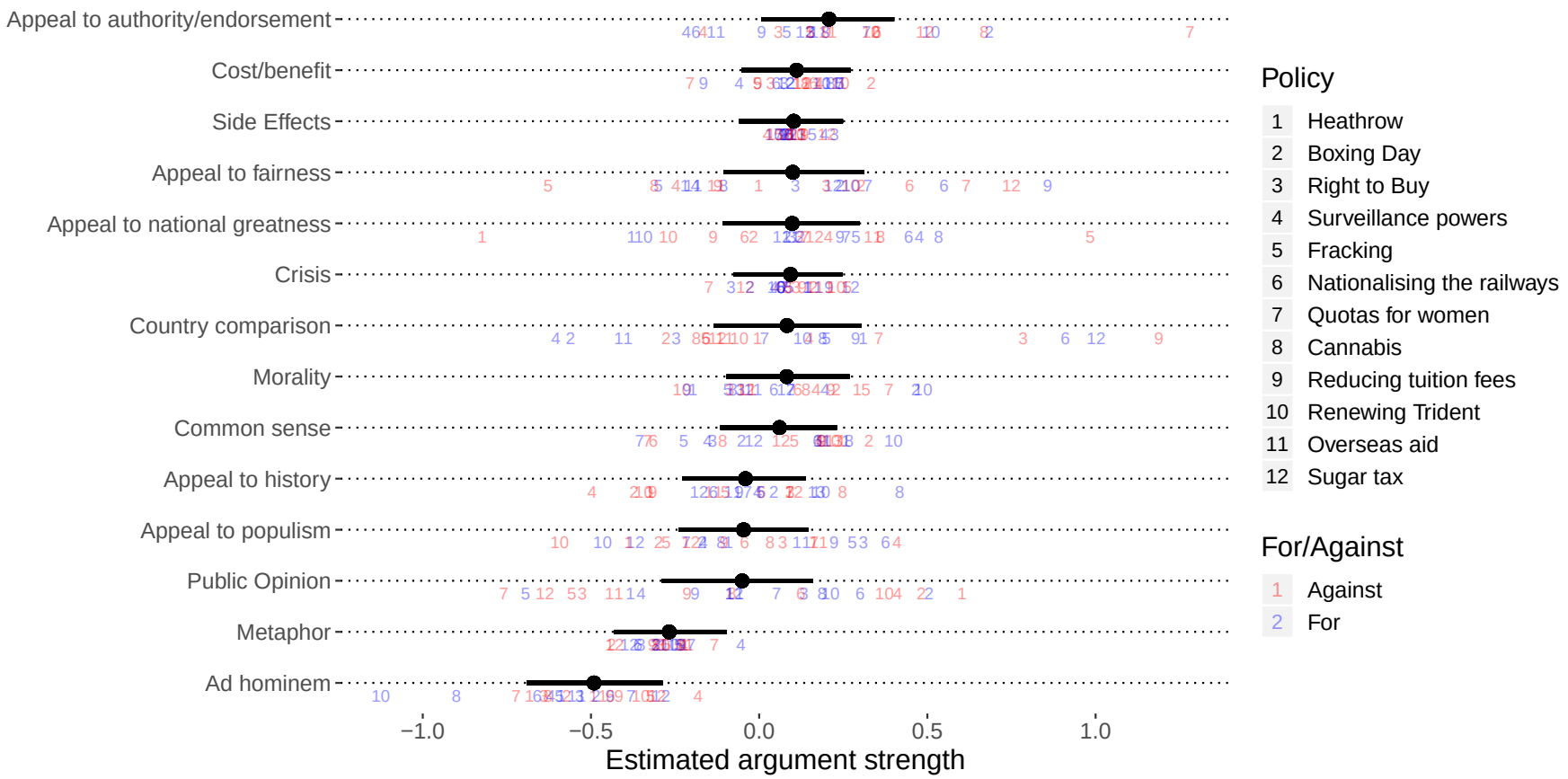


Figure 2: Estimated argument strengths

Two main patterns of interest arise in Figure 2. First, there is some variation in the average persuasive power of our 14 rhetorical elements. The estimates suggest that respondents have a clear aversion to arguments based on ad hominem attacks that impugn the character or motive of those on the opposite side of the issue as well as to arguments that are based on metaphor and imagery. The differences between the other element types are more modest, with at most suggestive evidence that arguments based on appeals to authority and appeals to fairness are most persuasive on average. Similarly, "crisis" oriented rhetorical appeals, and arguments which focus on relating policy issues to claims of national greatness are also, on average, modestly preferred by our respondents. Figure 3 shows that we can be reasonably confident that the average persuasive power of some of the element types is greater than others. For example, our posterior probability that appeals to authority of the kind that we tested 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.

Second, Figure 2 shows that there is 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 amongst 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 amonst the weakest that we include in the experiment. Similarly, while appealing to national greatness to oppose Fracking in the UK is relatively persuasive way to argue, opposing the expansion of Heathrow using similar appeals is not.[13] Argument strength heterogeneity is not equal across all element-types. For instance, the "metaphor" arguments all 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

[13] Compare "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." to "Great nations don't waste money on vanity projects, and the expansion of Heathrow would be nothing more than a project of national vanity."

be that we, or the MPs whose statements we adapted, are simply bad at ad hominem attacks but that such attacks can be effective when deployed more competently. It is also possible that certain types of argument are more or less amenable to some policy areas than others. 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 make more general claims about the 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.

Figure 4 shows 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. These appear to reflect the relative popularity of the policies themselves: nationalising the railways currently polls very well in the UK while fracking is very unpopular.

Controlling for Argument-Level Confounders

We have argued that by basing our experiment on multiple implementations of our latent concepts of interest, we have reduced the degree to which our analysis is likely to be subject to the type of confounding described in Grimmer and Fong (2019). As they argue, implementing multiple texts per treatment of interest helps to account for confounding by other correlated text features by allowing the analyst to average over those features. However, it remains the case that this approach will only help if the variation in background features is uncorrelated with our latent treatments. When this is not the case, marginalising over the omitted features will not recover unbiased estimates of the causal effects of interest.

[14] See, for example, work studying analogous rhetoric and the specific effects it has on attitudes towards public spending by Barnes and Hicks (2019), as well as Feldman and Hart (2016), who provide evidence about the motivational effects of political efficacy messages in the specific context of climate-related policy.

	authority	cost/benefit	side effects	fairness	national greatness	crisis	country comparison	morality	common sense	history	populism	public opinion	metaphor	ad hominem
authority		0.83	0.85	0.8	0.8	0.87	0.82	0.86	0.91	0.98	0.99	0.98	1	1
cost/benefit	0.17		0.55	0.54	0.54	0.58	0.61	0.6	0.72	0.95	0.94	0.92	1	1
side effects	0.15	0.45		0.55	0.5	0.56	0.59	0.6	0.7	0.95	0.95	0.91	1	1
fairness	0.2	0.46	0.45		0.51	0.53	0.57	0.56	0.64	0.89	0.89	0.86	1	1
national greatness	0.2	0.46	0.5	0.49		0.53	0.54	0.56	0.64	0.88	0.89	0.87	1	1
crisis	0.13	0.42	0.44	0.47	0.47		0.56	0.55	0.66	0.93	0.93	0.89	1	1
country comparison	0.18	0.39	0.41	0.43	0.46	0.44		0.5	0.55	0.85	0.85	0.81	1	1
morality	0.14	0.4	0.4	0.44	0.44	0.45	0.5		0.59	0.9	0.87	0.86	1	1
common sense	0.09	0.28	0.3	0.36	0.36	0.34	0.45	0.41		0.85	0.85	0.82	1	1
history	0.02	0.05	0.05	0.11	0.12	0.07	0.15	0.1	0.15		0.52	0.54	1	1
populism	0.01	0.06	0.05	0.11	0.11	0.07	0.15	0.13	0.15	0.48		0.51	0.99	1
public opinion	0.02	0.08	0.09	0.14	0.13	0.11	0.19	0.14	0.18	0.46	0.49		0.97	1
metaphor	0	0	0	0	0	0	0	0	0	0	0.01	0.03		0.99
ad hominem	0	0	0	0	0	0	0	0	0	0	0	0	0.01	

Figure 3: 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)

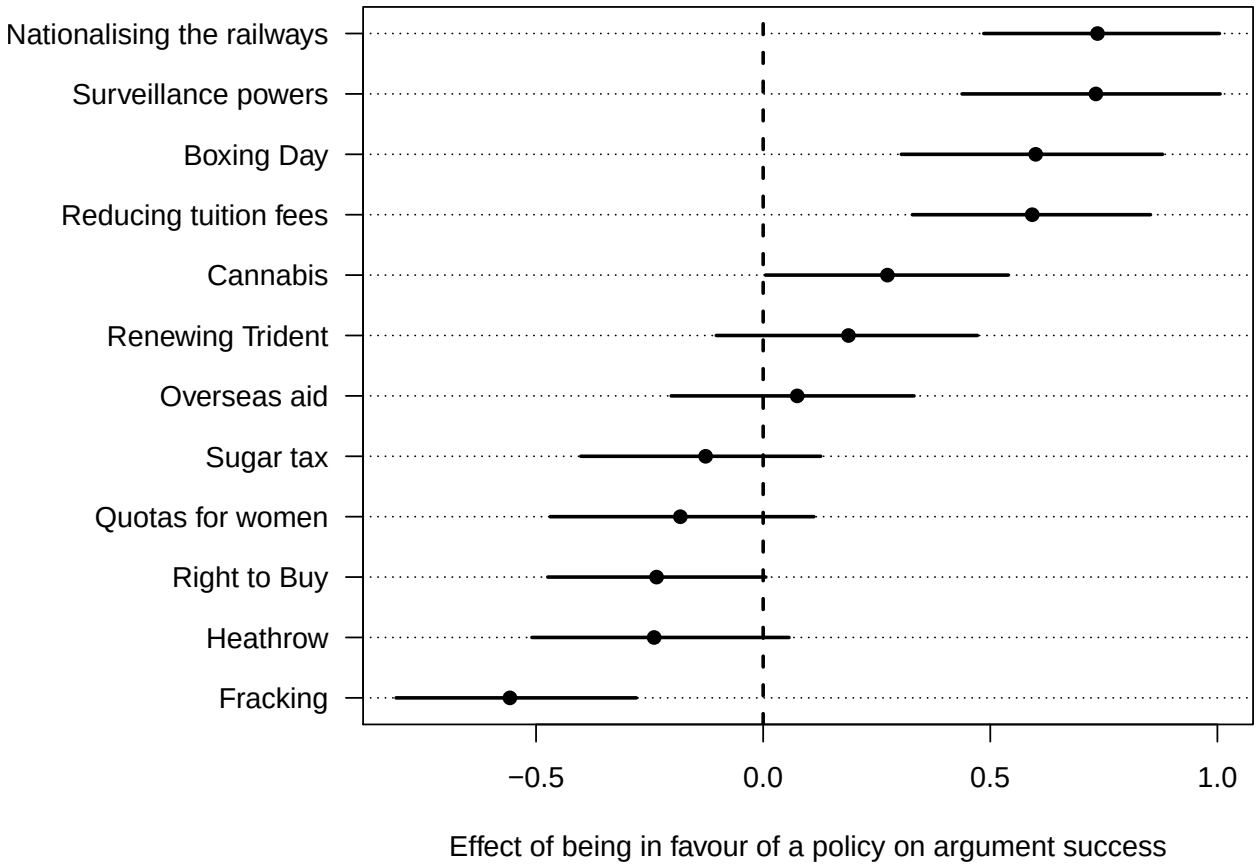


Figure 4: Policy effects

Our design and modelling strategy facilitate further ways to 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 simply control for potential confounding features when estimating the element effects. In the following analysis, we expand the second stage of our model to incorporate features of our arguments that may correlate with the rhetorical styles and which we can measure directly from the argument texts. In particular, we adapt equation 3 to include a vector of K argument-level measures, which we denote $x_{k,j}$:

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

$$\mu_e \sim N(0, \omega) \tag{5}$$

$$\nu_j \sim N(0, \sigma_{e(j)}) \tag{6}$$

The parameters γ_k represent average linear effects of text-feature k on argument strength. To illustrate the application of this control strategy, we have identified six argument-level variables which, in each case, represent features of our argument texts that might plausibly confound the effects of our rhetorical styles. We include argument length;[15] argument readability;[16] positive and negative tone;[17] overall emotional language;[18]; and fact-based language.[19]

Figure 5 presents the results from this model, and compares them with the estimates from equation 3 above. The left-hand panel of the plot shows the standardised posterior point estimates and intervals of the γ_k parameters from equation 5, and the right-hand panel compares the point estimates of the element average parameters $\mu_{e(j)}$ from the models with and without controls.

Of the 6 argument-level control variables we include, only length has a clearly significant ef-

[15] Measured in number of words.

[16] Measured using Flesch's Reading Ease Score (Flesch, 1948)

[17] Measured as the proportion of words in each argument listed in the positive and negative categories of the Affective Norms for English Words dictionary (Nielsen, 2011)

[18] Measured as the proportion of words in each argument listed in the 'affect' category from the 2015 Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker, Francis and Booth, 2001))

[19] Measured as the proportion of words in each argument listed in the 'quantitative' and 'numeric' categories of the LIWC dictionary.

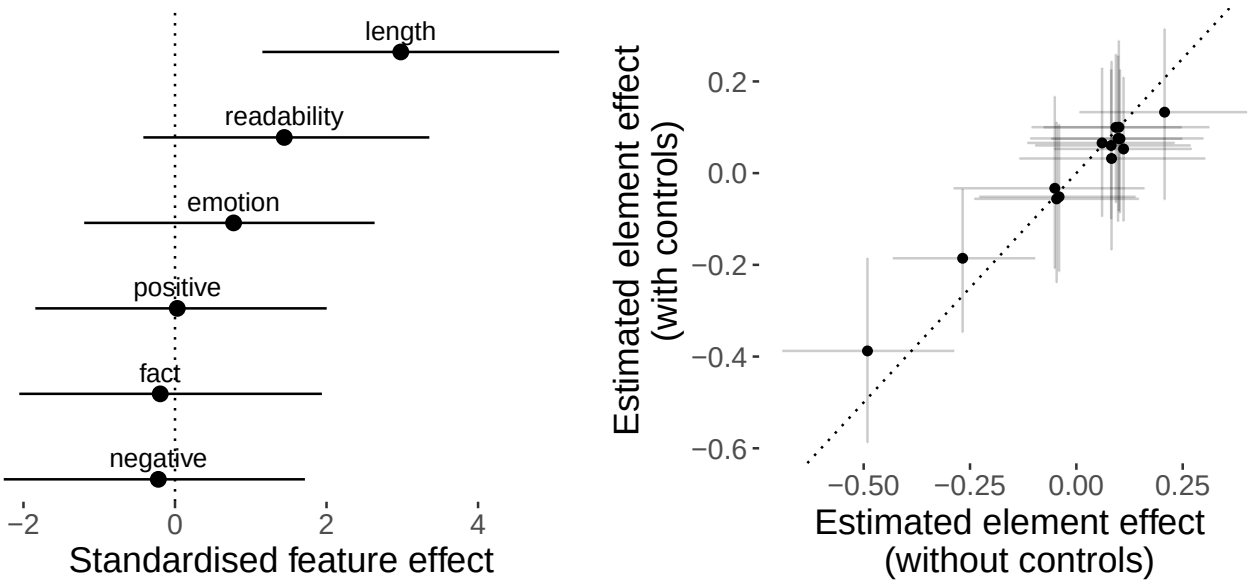


Figure 5: Control variable coefficient estimates (left) and comparison of element average effects with and without controls (right)

fect on argument persuasiveness. The estimates in the left-hand panel suggest that, on average, respondents find arguments with more words somewhat more persuasive than arguments with fewer words. We also find some suggestion that readability and emotional content positively predict persuasiveness. The right-hand panel of the figure also demonstrates that, in this particular case, controlling for the additional text features has only limited consequences for the rhetorical style effects. There is a slight attenuation of the differences between the rhetorical styles because the two least popular element types did have arguments that were somewhat shorter than average.

In this instance controlling for these variables does not materially change the overall conclusions regarding which element types are more or less effective on average. To reiterate, a common concern with survey experiments is that the latent features of interest to researchers may correlate with other unintended variation in the treatment texts that scholars construct. If those unintended features also affect a study's outcome variable, then the estimated effects of the latent treatments of interest will be biased. Although we find only limited evidence of such confounding here, a major strength of our design is that it gives researchers the ability to assess the robustness of the effects of their latent treatments to other potentially confounding features of this sort, after the experiment is completed. This is only possible because we combine many

instances of each latent feature with a modelling approach which allows us to control for other observable features of the treatment texts.

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.[20] At the same time, the significant 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 the aggregate, 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 "variance component" model that describes the extent of shared versus differential responsiveness to the treatments as a function of covariates. This model describes the strength of each argument for a person with a given set of covariate values as the sum of a set of random effects ξ_{jl}, each of which applies to individuals i as a function of their characteristics X_{il} on covariate l:

$$\begin{aligned} \alpha_{ij} &= \delta_{p(j),s(j)} + \nu_{ij} \\ \nu_{ij} &= \sum_{l} \xi_{jl} X_{il} \\ \xi_{jl} &= N(0, \sigma_l) \end{aligned} \tag{7}$$

We use dummy variables $X_{il} \in 0, 1$ in this model, so each variance component l corresponds to the magnitude of the idiosyncratic treatment effects that apply only to respondents with that characteristic. A small variance component associated with a covariate implies that individuals with that characteristic respond similarly to the full set of arguments, all else equal. A large variance component associated with a covariate implies that respondents with that covariate

[20] In appendix figure 8 we show results from a model where we allow the effects of each of our rhetorical elements to vary by demographic groups. Although we find some differences that are consistent with general intuitions about UK politics, the differential effects are small in size and imprecisely estimated.

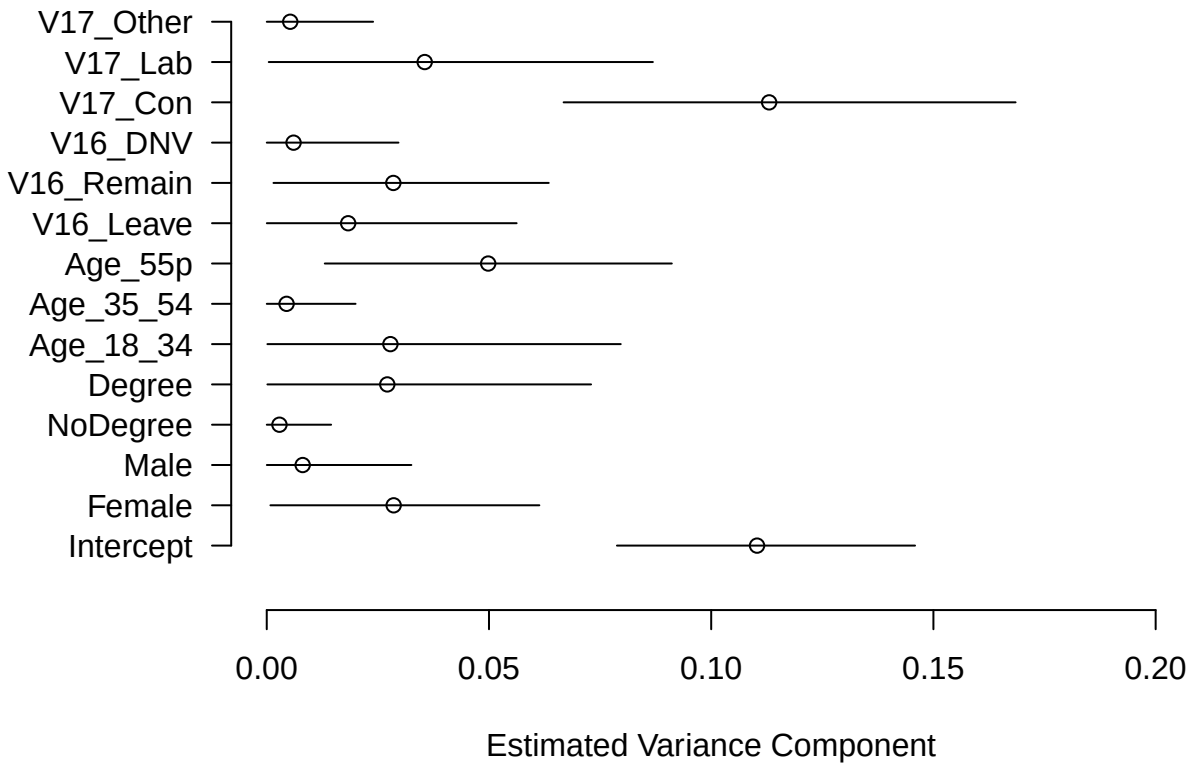


Figure 6: Estimated variance components associated with different respondent demographic attributes.

find arguments differentially persuasive from other respondents.[21] Given the estimates from this model, it is possible to recover the correlation in the argument strengths α_{ij} for any pair of individuals i with different covariate values X_{il}. When these correlations are large, it implies that individuals with those characteristics respond similarly to arguments in general, when they are low, it implies that such individuals find different arguments persuasive.

Figure 6 shows the estimated values of σ_l^2, which correspond to the relative contributions of the different variance components to variation in the persuasiveness of arguments across respondents with different characteristics. The relatively large value for the intercept indicates that there is a substantial common component to persuasiveness that applies to all individuals regardless of demographics. The largest component we estimate is for voting Conservative in 2017, indicating that this is the most important demographic characteristic for explaining variation in

[21]The additive functional form of the variance component model prevents negative correlations, so this model does not allow for the possibility that one group might systematically find relatively unpersuasive the arguments that another group found relatively persuasive. We have used the bivariate "correlated strength" model presented in the next section to confirm that this is not the case and that this functional form restriction is not important to the results.

argument persuasiveness. If we compare a respondent who is male, without a university degree, over 55 years old, who voted Leave and who voted Labour to an otherwise identical profile for a Conservative voter, the estimated correlation in their argument strength estimates α_{ij} is 0.73. This is still a reasonably high correlation: even individuals who vote for different parties largely agree on which of the arguments we presented are relatively strong. If we instead compare the aforementioned Leave-Labour voter to a Remain-Labour voter with the same characteristics, the estimated correlation in their argument strength estimates α_{ij} is 0.92.

If we take an extreme comparison, between a female, with university degree, 18-34 years old, who voted Remain and Labour versus a male, without a university degree, over 55 years old, who voted Leave and Conservative, the correlation declines to 0.40. While this is no longer a high correlation, it is also not especially close to zero. Thus, there is evidence that political differences shape which arguments respondents find persuasive, but this heterogeneity seems to be moderate in magnitude. Perhaps surprisingly, it is more oriented around the parties that voters supported in 2017 than the increasingly salient Leave vs Remain dimension of conflict in UK politics.

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 use a "correlated strength" model to assess the possibility that arguments might have different relative strengths when facing arguments on the same side of the issue than they do when facing arguments on the opposite side of the issue. These strengths are assumed to be correlated with unknown correlation coefficient ρ, and our goal is to estimate the correlation. This correlation tells us whether the relative strengths of the arguments, across our entire experiment, tend to be the same in the same-side comparisons as the opposite-side comparisons (large ρ, near 1) or whether they tend to be unrelated (small ρ, near 0). Where $k = 1$ when the argument is presented against an argument on the opposite side and $k = 2$ where the argument

is presented against an argument on the same side:

$$\alpha_{jk} = \delta_{p(j)} + \nu_{jk}$$
$$\nu_j \sim MVN(0, \Sigma) \tag{8}$$

Σ is 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: $\rho = 0.81$ (0.69-0.91), $\sigma_1 = 0.33$ (0.29-0.38), 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.

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.

Conclusion

We have described a new experimental design and modelling strategy for testing the relative persuasiveness of different types of political arguments. Basing our design on the types of rhetoric that are regularly found in real-world political speeches in the UK, we implemented an experiment using 336 individual arugments pertaining to 14 rhetorical elements and 12 policy issues. Combining a Bradley-Terry style model with a series of hierarchical models, we have demonstrated (1) that there are modest differences in the persuasive power of different rhetorical elements, with "appeals to authority" and "cost/benefit" amongst the strongest types we test, and "ad hominem" and "metaphor" amongst the weakest; (2) that there is significant heterogeneity in argument strength *within* element types, implying external validity concerns for existing studies that rely on single implementations of latent treatments in texts; (3) that evaluations of persuasiveness based on comparisons of arguments on the same-side of an issue are highly positively correlated with those based on opposite-side comparisons; and (4) that there is a large degree of correlation in the perceived strength of political arguments across different types of respondent.

Although we based our experiment on rhetorical types that occur "in the wild" in British politics, future work could use our general approach of multi-instance treaments to distinguish between other rhetorical typologies of interest to social scientists. For example, there is ongoing debate over whether elites influence citizens by providing them with new information in their communications, or by subtly framing policy debates in a way that shifts the weights that citizens attribute to different evaluation criteria. A key difficulty in this literature has been establishing methodological approaches that allow researchers to distinguish between "framing" and "information" treatments (Leeper and Slothuus, 2018). Our approach of including multiple instances of each type of treatment could help to lessen concerns about confounding between these concepts.

Similarly, studies of classical rhetoric identify three distinct elements of persuasive speech: ethos, which refers to appeals to the speaker's character or credibility; logos, which refers to appeals to reason and logic; and pathos, which refers to appeals to audience emotions. These three means of persuasion form the core of Aristotle's *Rhetoric* (Aristotle, c.322 BCE), and have

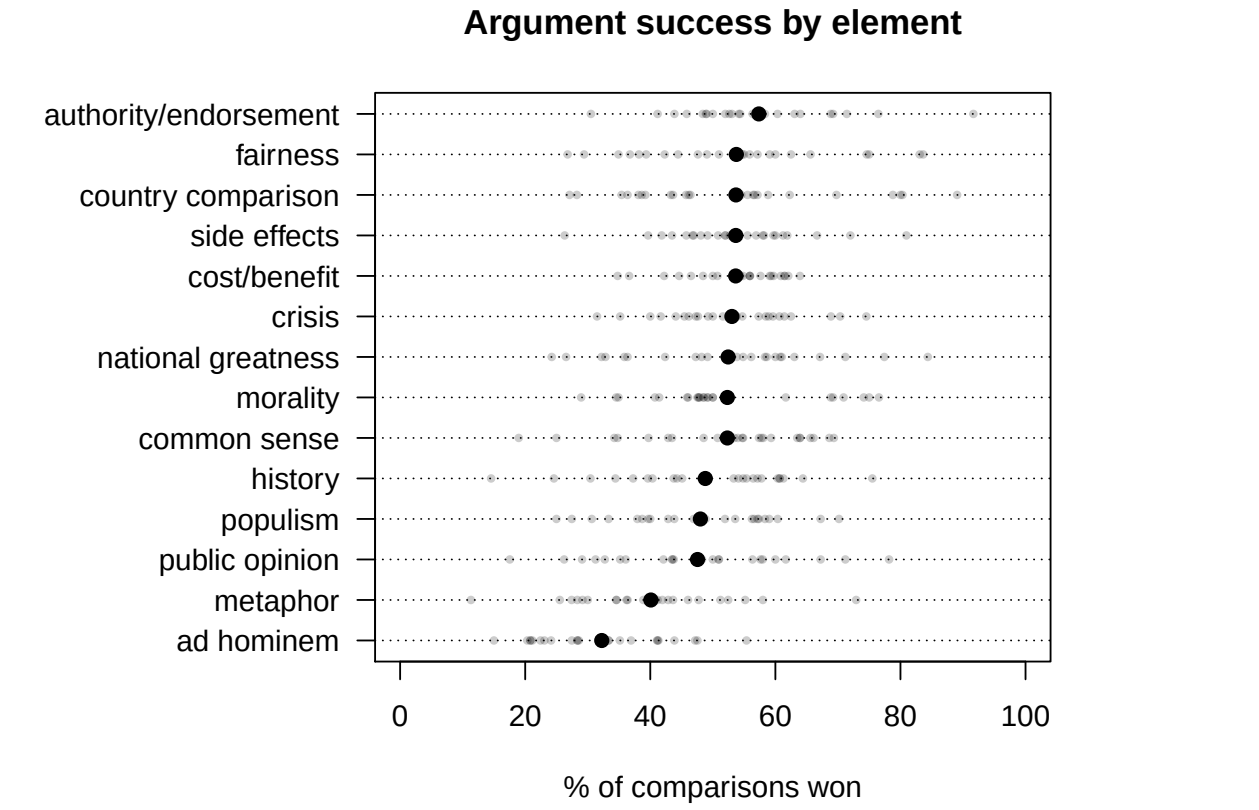
continued to inspire thought into the uses of rhetoric and persuasive speech in contemporary research (Charteris-Black, 2011). The experimental design we introduce here could likewise be used to measure the importance of these forms of argument and their relative persuasive power.

Finally, we have focussed on measuring persuasiveness as it relates to the approval of specific policies, but our framework could also be straightforwardly applied to questions regarding political advertising in election campaigns (Gerber et al., 2007), or messaging strategies intended to increase voter participation rates (Gerber and Green, 2000) or donations to charity (DellaVigna, List and Malmendier, 2012), all of which are questions that have featured prominently in recent literature.

Appendix

Table 2: Policy issues

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



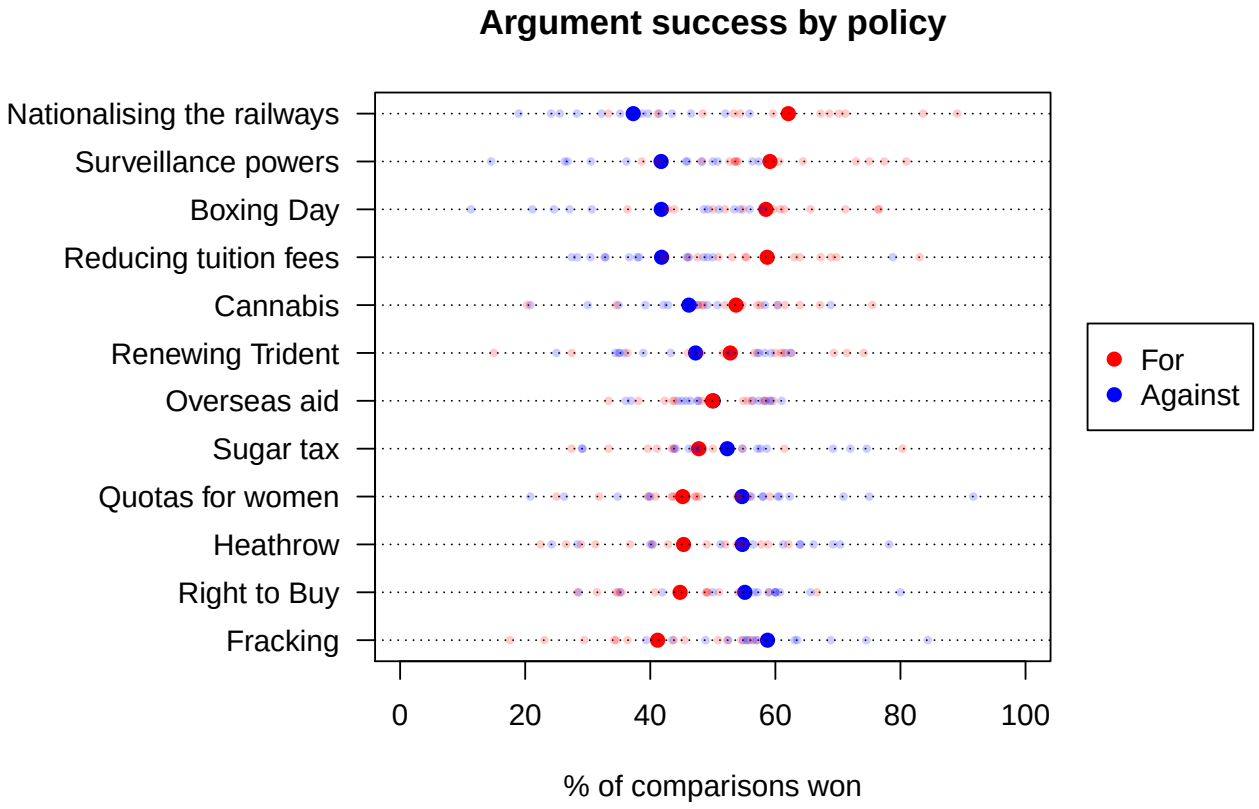


Figure 7: Raw data

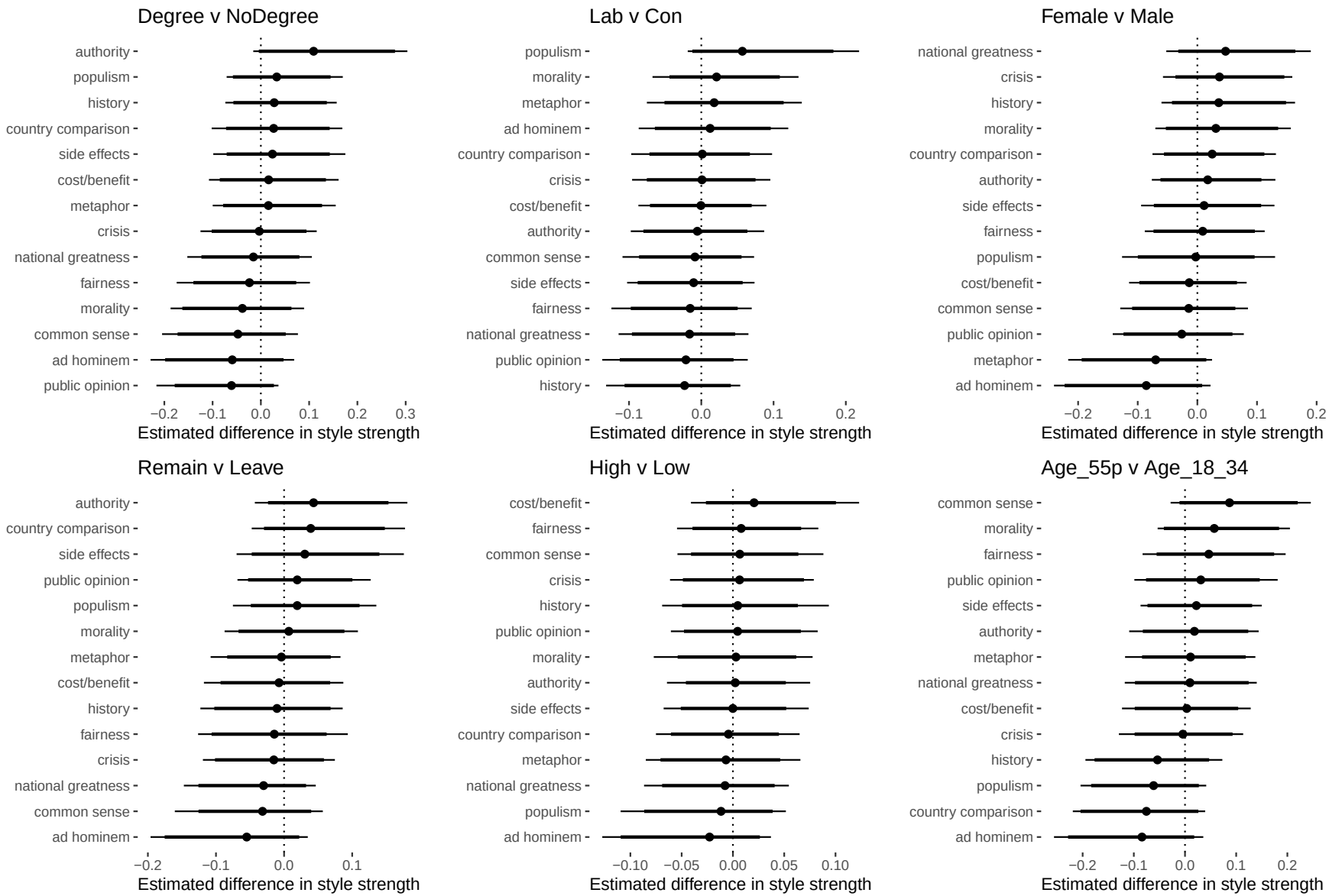


Figure 8: Difference in rhetorical element strength by covariate

Table 3: 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.
	For	Air travel should not just be for elites: expanding Heathrow will make flights cheaper and more accessible for ordinary working people.

Table 3: Treatment texts – Building a third runway at Heathrow *(continued)*

Element	For/Against	Text
Appeal to populism	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.
Morality	For	It is wrong to let local interests get in the way of the national interest in having an internationally competitive airport.
	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.
	For	A 2014 poll showed that many more people were in favour of a third runway at Heathrow than opposed it.

Table 3: Treatment texts – Building a third runway at Heathrow *(continued)*

Element	For/Against	Text
Public Opinion	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 4: 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.
	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.

Table 4: Treatment texts – Closing large retail stores on Boxing Day *(continued)*

Element	For/Against	Text
Cost/benefit	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.
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 5: 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.

Table 5: 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.

Table 5: Treatment texts – Extending the Right to Buy *(continued)*

Element	For/Against	Text
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 6: 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.
Appeal to fairness	For	Extending surveillance powers is the fairest way to protect all people in our country.
	Against	Excessive surveillance 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.

Table 6: Treatment texts – Extension of surveillance powers in the UK *(continued)*

Element	For/Against	Text
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.
	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.

Table 6: Treatment texts – Extension of surveillance powers in the UK *(continued)*

Element	For/Against	Text
Public Opinion	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.

Table 7: Treatment texts – Fracking in the UK

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.

Table 7: Treatment texts – Fracking in the UK *(continued)*

Element	For/Against	Text
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.
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.

Table 7: Treatment texts – Fracking in the UK *(continued)*

Element	For/Against	Text
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 8: 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.

Table 8: Treatment texts – Nationalisation of the railways in the UK *(continued)*

Element	For/Against	Text
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.
Crisis	For	The current franchise system has been responsible for a series of transport crises throughout the country.
	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.
	For	Poll after poll shows that people are dissatisfied with the train services they receive throughout the country.

Table 8: Treatment texts – Nationalisation of the railways in the UK *(continued)*

Element	For/Against	Text
Public Opinion	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 9: 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.
Appeal to populism	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.
	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.

Table 9: Treatment texts – Quotas for women on corporate boards *(continued)*

Element	For/Against	Text
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.

Table 9: Treatment texts – Quotas for women on corporate boards *(continued)*

Element	For/Against	Text
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 10: 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.
Appeal to authority	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.
	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.
	For	It is common sense to legalise cannabis given that it is less dangerous than legal drugs like alcohol and tobacco.

Table 10: Treatment texts – Reducing the legal restrictions on cannabis use *(continued)*

Element	For/Against	Text
Common sense	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.
	For	One attractive side-effect of legalising canabis is the additional tax revenue it would bring to the government.

Table 10: Treatment texts – Reducing the legal restrictions on cannabis use *(continued)*

Element	For/Against	Text
Side Effects	Against	An unintended consequence of legalizing cannabis would be to encourage children to try it.

Table 11: 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?
	For	The existing tuition fee system is not working as planned, and so it is simply common sense to make changes.

Table 11: Treatment texts – Reducing university tuition fees *(continued)*

Element	For/Against	Text
Common sense	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.
Morality	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."
	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.

Table 11: Treatment texts – Reducing university tuition fees *(continued)*

Element	For/Against	Text
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 12: 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.

Table 12: Treatment texts – Renewing Trident *(continued)*

Element	For/Against	Text
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.
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 deterrant 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.

Table 12: Treatment texts – Renewing Trident *(continued)*

Element	For/Against	Text
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 13: 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?"
	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.

Table 13: Treatment texts – Spending 0.7 per cent of GDP on overseas aid *(continued)*

Element	For/Against	Text
Common sense	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.
	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.

Table 13: Treatment texts – Spending 0.7 per cent of GDP on overseas aid *(continued)*

Element	For/Against	Text
Side Effects	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 14: 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.
Appeal to history	For	There is plenty of historical evidence that Government action can change people's behaviour, and a sugar tax would be no different.
	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.
	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.

Table 14: Treatment texts – Sugar tax in the UK *(continued)*

Element	For/Against	Text
Common sense	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.
	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.

Table 14: Treatment texts – Sugar tax in the UK *(continued)*

Element	For/Against	Text
Side Effects	Against	A likely unintended consequence of a sugar tax would be for companies to introduce other unhealthy sweetening agents into our drinks.

References

Agresti, Alan. 1992. "Analysis of ordinal paired comparison data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 41(2):287–297.

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.

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

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.

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.

Davidson, Roger R. 1970. "On extending the Bradley-Terry model to accommodate ties in paired comparison experiments." *Journal of the American Statistical Association* 65(329):317–328.

DellaVigna, Stefano, John A List and Ulrike Malmendier. 2012. "Testing for altruism and social pressure in charitable giving." *The quarterly journal of economics* 127(1):1–56.

DellaVigna, Stefano and Matthew Gentzkow. 2010. "Persuasion: empirical evidence." *Annu. Rev. Econ.* 2(1):643–669.

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

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.

Fong, Christian and Justin Grimmer. 2016. Discovery of treatments from text corpora. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. pp. 1600–1609.

Gerber, Alan, James G Gimpel, Donald P Green and Daron R Shaw. 2007. "The influence of television and radio advertising on candidate evaluations: Results from a large scale randomized experiment." *Unpublished paper, Yale University* .

Gerber, Alan S and Donald P Green. 2000. "The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment." *American political science review* 94(3):653–663.

Glenn, WA and HA David. 1960. "Ties in paired-comparison experiments using a modified Thurstone-Mosteller model." *Biometrics* 16(1):86–109.

Graham, Matthew and Alexander Coppock. 2019. "The Perils of Self-Assessed Attitude Change." *Working Paper* .

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

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.

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* .

Kant, Immanuel. 1790. *Critique of the Power of Judgment*. Cambridge University Press, 2001.

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.

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.

Madison, James. 1788. 63. In *The Federalist Papers*. Oxford University Press, 2008.

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

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

O'neill, John. 2002. "The rhetoric of deliberation: Some problems in Kantian theories of deliberative democracy." *Res Publica* 8(3):249–268.

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.

Rao, PV and Lawrence L Kupper. 1967. "Ties in paired-comparison experiments: A generalization of the Bradley-Terry model." *Journal of the American Statistical Association* 62(317):194–204.

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.

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

Tutz, Gerhard. 1986. "Bradley-Terry-Luce models with an ordered response." *Journal of mathematical psychology* 30(3):306–316.