A Double Standard? Gender Bias in Voters' Perceptions of Political Arguments

Lotte Hargrave *University College London*

Do the styles politicians use influence how voters evaluate them, and does this matter more for women than for men? Politicians regularly use anecdotal arguments, emotional appeals, and aggressive attacks when communicating with voters. However, that women politicians have been branded as 'nasty', 'inhuman', and 'unfeminine' suggests these strategies may come at a price for some. I report on a novel survey experiment assessing whether voters are biased in their perceptions and evaluations of politicians' communication styles. By manipulating politician gender and argument style I assess, first, whether politicians incur backlash when violating gender-based stereotypes, and second, whether differential perceptions of the styles themselves explain this backlash. I find that style usage has important consequences for how voters evaluate politicians, but that this is not gendered. These results have important implications as they suggest that women politicians may not need to conform to stereotype-expected behaviours to receive positive voter evaluations.

Keywords: gender; survey experiment; stereotypes; style; voting behaviour

Word count: 9,996

Supplementary Material

Online appendixes are available at [JOURNAL TO INSERT LINK ONCE READY].

Data Availability Statement

Replication data for this article can be found in Harvard Dataverse at [INSERT ONCE READY].

Acknowledgements

I am incredibly grateful to Lucy Barnes, Jack Blumenau, Ana Catalano Weeks, Jennifer Hudson, Kasim Khorasanee, Markus Kollberg, Hannah McHugh, Rebecca McKee, Alice Moore, Tom O'Grady, Meg Russell, Jess Smith, Sigrid Weber, and Chris Wratil for helpful conversations and insightful comments on, in some cases, numerous drafts of the design, pre-analysis plan, and paper. I also thank participants at the University of Bath PoLIS Department Research Seminar, EPSA 2022, and at various working group seminars at University College London for their thoughtful feedback on this work.

Financial Support

None.

Competing Interets

None.

Introduction

Discussion and debate lie at the heart of democratic politics. When communicating with the public, politicians make use of a wide range of argumentation strategies such as empathetic appeals, emotional arguments, and aggressive attacks. What politicians say, and importantly, how they say it, matters a great deal. While politicians expressing empathy, emotion, and aggression is unremarkable, certain politicians may face harsher penalties than others when they express, or do not express, these styles. Former UK Prime Minister Theresa May was known for a style that was characterised as unemotional, unempathetic, and robotic. May's style received significant press commentary with some asking "what is wrong" with her, others branding her as "inhuman and uncaring", or suggesting her "inability to show emotion to the public proves that she isn't fit to be Prime Minister". In the eyes of the public too, May was seen as having a "cold personality". This led to her infamous characterisation in the press and across social media as "the Maybot". Yet, when her resignation speech was uncharacteristically marked with emotion and tears, it prompted some to bid farewell to a leader who was "almost human after all".

Theresa May's style is one that goes against what is stereotypically expected of women in politics (Schneider and Bos, 2019), and she has been branded as inhuman, cold, and robotic as a result. While this anecdote illustrates only one example of how the styles of women are evaluated, there is ample evidence that women politicians more broadly incur penalties when they do not conform to stereotype-congruent behaviours (Bauer, 2015*a*; Boussalis et al., 2021; Cassese and Holman, 2018). Men politicians, by contrast, have been shown to be successful both when they conform to stereotype expected behaviours – such as being ambitious (Saha and Weeks, 2022) – *and* when they subvert expectations – such as being emotional or communal (Gleason, 2020; Okimoto and

Brescoll, 2010). Indeed, past work has shown there to be clear evidence of asymmetric standards in how men and women politicians are evaluated.

In this paper, I assess whether voters are biased in their perceptions and evaluations of the ways in which politicians communicate, and, consequently, whether voters' evaluations of elite political communication are gendered. While there is a robust literature on the influence of gender stereotypes on voter evaluations in politics more broadly, so far scholars have neglected to unpack whether voters differentially *perceive* politicians' behaviour based on gender alone. I make an important contribution by being the first to address this question, and to assess whether differential perception of styles themselves may serve as a potential mechanism through which voters' gendered evaluations of politicians become manifest.

To test these questions, I focus on styles for which the gender stereotypes literature has outlined clear expectations about men and women's behaviour. Work to date examining how voters evaluate stereotype-(in)congruent behaviour has focused on isolated traits (e.g., Bauer, 2015*a*; Cassese and Holman, 2018). Instead, while I analyse the effect of each style separately, I make progress by focusing on how voters evaluate politicians' use of a diversity of styles that are consistent with both feminine "communal" and masculine "agentic" stereotypes (Schneider and Bos, 2019). By doing so, I assess whether there are asymmetric standards in the degree to which men *and* women politicians are punished for stereotype-incongruent behaviours. I focus on the use of emotion, aggression, and evidence (both drawing on *statistical* and *anecdotal* evidence). In a novel survey experiment, I present UK voters with speeches where the argument style and the gender of the politician delivering the argument are varied. Through these manipulations I assess, first, whether politicians incur negative evaluations from voters when they deploy styles that are stereotype-incongruent, and second, whether voters' differential perception of styles themselves might explain the backlash politicians receive.

I report four main findings. First, I find that politicians' style usage has important consequences for how voters evaluate them. Politicians are liked more when they are emotional or draw on anecdotes, however, they are regarded as more competent when they are unemotional and make use of statistical evidence. Second, contrary to expectations from the stereotypes literature, I find no evidence that these evaluations are gendered. That is, while all politicians are perceived as less likeable when they are aggressive, there is no evidence that women in particular incur negative evaluations. Third, while there is clear evidence that voters can identify the styles politicians use, I also find no evidence that voters' perceptions of the styles themselves are gendered. Voters do not perceive unemotional arguments by women as any less emotional than unemotional arguments by men, nor do they perceive aggressive arguments by women to be any more aggressive than equally aggressive arguments by men. Fourth, I find some evidence that these evaluations differ by voter gender. Women voters reward women politicians for stereotype-congruent behaviour: they give a larger likeability and competence reward to women who are emotional and perceive arguments by women to be more emotional.

The main finding I document is, therefore, that there is little evidence of gender bias in the forum I study. These findings do not, however, imply that there is no gender bias at play in politics. I study voter evaluations of the likeability and competence of politicians as gendered and non-gendered accounts of leadership alike suggest they are some of the key qualities expected of politicians (Clarke et al., 2018; Huddy and Terkildsen, 1993). However, even if voters equally evaluate men and women for the styles they use, I do not assess how likeability and competence may have downstream consequences for voting. Previous US-based work has suggested that competence evaluations may be more important in voters' evaluations of women (Ditonto, 2017), and therefore women may *need* to be perceived as more competent than men to get elected. The media may also play an important role in the *framing* of women's behaviour. If the media frame women's

behaviour in a more negative light than men's this may in turn feed into how voters evaluate politicians even if voters' direct judgements are not themselves gendered. Finally, my results suggest that voters deem politicians to be less competent when they express styles traditionally associated with feminine "communal" stereotypes (Schneider and Bos, 2019). Therefore, if the distribution of actual style usage is different for men and women then this may still lead to bias. If it is true that women happen to make greater use of communal styles – such as emotion or anecdotes – than men, then they may incur negative competence evaluations simply because of the styles they use. Evidence from a variety of contexts has shown that women politicians are more emotional (Dietrich, Hayes and O'Brien, 2019) and draw more on anecdotes (Hargrave and Langengen, 2021), although recent work has shown that women have made decreasing use of these styles over time in the UK (Hargrave and Blumenau, 2022). In short, identifying the mechanism through which biases emerge in the judgement of politicians' behaviour is challenging, and I make an important contribution by shedding light on one crucial aspect of this complex process: whether voters' perceptions of and attitudes towards elite political communication are gendered.

A commonly held perception is that men's leadership styles are preferred for political office than women's (Fox and Oxley, 2003), and that when women try and adopt these styles, they violate feminine stereotypes, and will lose out. However, the findings I present here suggest the reality for women may be more positive than this, as the expectation that they must avoid these styles for fear of negative backlash from voters is somewhat misguided. Recent work has shown that discrimination can occur when individuals perceive that others are likely to discriminate and that voters tend to overestimate the degree to which *others* are biased (Bateson, 2020). Documenting that voters may not be as biased towards stereotype-incongruent women is important, then, as it may potentially help to further reduce voter bias. While of course women may be sanc-

tioned for stereotype-incongruent behaviour from other sources, the findings I present here suggest that voters do not punish women in the way common theories of gender stereotyping may have predicted.

Gender, stereotypes, and voter backlash

Why might voters hold different expectations for how men and women politicians behave? Gender role theory, a prominent psychological approach for explaining gender-based behavioural differences (Eagly and Karau, 2002), suggests that stereotypes concerning the typical behaviours of men and women lead to strong expectations for how individuals of different genders will and should behave. The social roots of such gender roles are thought to emerge from a type of statistical profiling: to the extent that, for diverse historical reasons, women and men display different behaviours on average, people internalise these patterns and the "corresponding attributes become stereotypic of women [men] and part of the female [male] gender role" (Eagly and Wood, 2012, 11). These behaviours are not always precisely defined; however, they are broadly divided into women's "communal" behaviours which are associated with emotionality, positivity, warmth, human interest, and caring for others, and men's "agentic" behaviours which are associated with aggression, logic, and leadership (Schneider and Bos, 2019).

Over time, the expectation that men and women should behave in ways associated with their gender is reinforced through a process of socialisation. Further, descriptive stereotypes, which purport to describe what group members *are* like, lead to prescriptive stereotypes, which purport to describe what group members *should* be like (Gill, 2004). The idea that women are, say, gentle and emotional eventually leads to the expectation that women should be gentle and emotional. Behaviour that is inconsistent with these stereotypes may then be punished via social sanctions imposed by others. The upshot is that gender roles become self-reinforcing, as societal beliefs about the typical

behavioural differences lead to ever-more entrenched patterns of gendered behaviour.

I focus on four *styles* for which the gender stereotypes literature outlines clear expectations for gendered usage. The first communal style is **emotion**. Women are expected to be more emotional (Dietrich, Hayes and O'Brien, 2019; Huddy and Terkildsen, 1993), and more positive specifically (Hargrave and Blumenau, 2022). The second communal style is **anecdotal evidence**. Women are thought to make greater use of anecdotes, which includes referencing personal experience, analogies, or stories (Childs, 2004; Hargrave and Langengen, 2021). The first agentic style is the use of **aggression**. Men are thought to be more aggressive (Grey, 2002), whereas women are thought to avoid this behaviour for fear that it is "negatively perceived by the electorate" (Childs, 2004, 190). The second agentic style is the use of **statistical evidence** (Mattei, 1998), which is linked to the idea that men are thought to be more "analytical, organised and impersonal" (Jamieson, 1995, 76).

How might the conformity to or violation of stereotype-congruent behaviours affect voter evaluations of politicians? Previous work has shown that women are subject to pressures to conform to stereotypes. When women express behaviours which are *consistent* with stereotypes of communality, they tend to be rewarded. Focusing on the behaviour of Angela Merkel, Boussalis et al. (2021) find Merkel is rewarded by voters for displays of happiness but punished for expressions of anger. Similarly, work by Gleason (2020) shows that women Supreme Court attorneys are successful when they use extensive emotional appeals. By contrast, when women express behaviours which are *inconsistent* with stereotypes of communality, they can incur backlash for it. For instance, Cassese and Holman (2018) find that women candidates in particular are vulnerable to attacks when they use negative campaigning. Although, work by Brooks (2011) finds that men and women candidates in the US are similarly penalised for their expressions of both anger and tears. How might these evaluations translate to men? Voters

have been shown to be less sensitive to men expressing styles which are incongruent with masculine "agentic" stereotypes (Gleason, 2020), and indeed studies suggest that men are successful when they both follow *and* subvert gendered expectations (Okimoto and Brescoll, 2010).

I seek to identify whether politicians' use of stereotype-(in)congruent styles leads to differential voter evaluations. Previous work has focused on a diversity of traits that might be influenced by conformity to stereotypes (Brooks, 2011; de Geus et al., 2021; Saha and Weeks, 2022), which can be divided along two lines. First, the extent to which politicians' conformity to stereotypes leads to voters feeling more *warmly* towards them as individuals. Women are expected to be communal, which includes being kind, caring, and compassionate (Eagly and Karau, 2002). Therefore, when they behave in ways that are instead consistent with agentic stereotypes, such as being aggressive or ambitious, they fall short of expectations about the communal qualities deemed appropriate for women (Schneider and Bos, 2019). Studies assessing voter evaluations of role-incongruent women confirm that women are judged less favourably by voters (Boussalis et al., 2021), whereas role-congruent women are regarded more favourably and rewarded via increased likeability evaluations (Bauer, 2017).

Second, conformity to stereotype-congruent behaviours may also influence the extent to which voters judge politicians' ability to perform their jobs to a high standard. Politicians are expected to be competent and, as the traditional occupants of political leadership positions, men's competence in these roles is assumed in a way that women's is not (Schneider and Bos, 2019). Recent experimental work has shown that women *overall* may actually have a small electoral advantage relative to men (Schwarz and Coppock, 2022) and that stereotypes about women's competence may have equalised in recent years (Donnelly et al., 2016). Nonetheless, a large literature has shown that women have historically been evaluated as less competent (Huddy and Terkildsen, 1993), and

that voters seek out more information about the qualifications of women than men (Ditonto, 2017). Consequently, I focus on both likeability and competence, which gendered and non-gendered accounts of leadership alike identify as some of the key qualities expected of politicians (Clarke et al., 2018; Huddy and Terkildsen, 1993). For both, I expect that women will be rewarded for expressing styles congruent with feminine "communal" stereotypes and punished for expressing styles congruent with masculine "agentic" stereotypes.

This study contributes to the literature discussed above in three key ways. First, work to date examining how voters evaluate stereotype-congruent or incongruent behaviour has focused on isolated behaviours, such as tears (Brooks, 2011) or negative attacks (Cassese and Holman, 2018). By contrast, while I analyse the effect of each style separately, I expand upon this to focus on how voters evaluate politicians' use of a diversity of styles that are consistent with both feminine "communal" and masculine "agentic" stereotypes. Therefore, I can identify whether women in particular incur penalties for stereotype-incongruent behaviour. Second, work on voter backlash has focused almost exclusively on the US. While European work has studied gender bias in voting behaviour, with few exceptions (Saha and Weeks, 2022) scholars have not assessed the influence *stereotypes* might have on UK voters' evaluations. Third, while there is a robust literature on the influence of stereotypes on voter *evaluations* in politics more broadly, so far scholars have neglected to unpack whether voters differentially *perceive* the behaviour of men and women, and whether differential perception of behaviour itself may in turn be responsible for how voters evaluate politicians. This study is the first to assess whether differential perceptions of styles themselves may serve as a potential mechanism through which voters' gendered evaluations of politicians might become manifest.

Differential perceptions of styles

What may explain why voters differentially penalise men and women for expressing styles that violate stereotypes? One mechanism through which these evaluations might become manifest is that – because of stereotypical expectations – voters simply perceive differently the styles that men and women use. That is, even in the absence of any *objective* differences in, say, the emotionality of a speech, voters "hear" that speech as more or less emotional depending on the politician's gender. This differential perception of style may then, in turn, lead to voters' differential evaluation of the likeability or competence of men and women. As emotion is a style that is congruent with feminine communal stereotypes. Therefore, when women are not emotional, this may violate expectations about the supposed emotional sensitivity of women and be of particular note to voters. Because women are seen as particularly unemotional, they, relative to men, may be ascribed more negative evaluations and be seen as particularly "cold and unlikeable".

Aggression is masculine stereotype-congruent. Therefore, when voters see a man delivering an aggressive speech, this may only serve to confirm pre-existing stereotypes. Stereotypical expectations dictate, however, that women are *not* expected to be aggressive. Voters may therefore pay more attention to aggressive women because this violates stereotypes. Because voters see a woman as particularly aggressive this may, in turn, affect their likeability or competence evaluations of that woman. The experiences of former US presidential candidate Hillary Clinton – who was described as "too angry, aggressive, and unfeminine" (Brooks, 2013, 1) – may help to further explain this dynamic. Clinton may have been perceived as more aggressive than her men counterparts even when she may not actually have behaved in such as way because aggression is congruent with masculine stereotypes. Therefore, because Clinton's use of aggression is particularly noticeable, this may feed into her being evaluated more negatively.

There may also be differences in the degree to which voters perceive men and women's use of anecdotes and statistics as evidence-based. While existing work has presented mixed conclusions about the persuasiveness of these evidence types, it is broadly concluded that statistical arguments help ensure argument credibility (Hornikx, 2018). As outlined above, studies have highlighted that men's competence is often assumed, while women's is not. Therefore, while I expect that arguments by *all* politicians will be perceived as more evidence-based when they contain statistics, arguments by women *in particular* will be perceived as more evidence-based when they include statistical evidence compared to anecdotal evidence. However, because of the assumed competence of men, the difference between anecdotal and statistical arguments delivered by men will be smaller.

Therefore, I assess whether voters perceive men and women's styles differently even in circumstances where there are no differences, and whether differential style perception may work as a mechanism for explaining variation in the likeability and competence evaluations that politicians receive.

Research Design

I test my expectations with a vignette survey experiment. In the experiment, respondents were tasked with reading an argument delivered by a fictitious politician. Respondents were randomly assigned to different treatment conditions where four attributes were randomised. First, the **style** of interest: *emotion*, *aggression*, or *evidence*. Second, the **treatment status** of the style. For emotion, this is the *control* condition which is a neutral, non-emotional speech and the *treatment* condition which is highly emotional. For aggression, this is the *control* condition which is a neutral, non-aggressive and the *treatment* condition which is highly aggressive. For evidence, the *control* condition is statistical evidence and the *treatment* condition is anecdotal evidence.

Third, the **policy area**: *housing*, *health*, or *transport*. While I am not interested in making inferences about specific policies, I include several as there may be a concern that any effects uncovered for, say, women and emotionality on health may not translate to another area such as transport. Work investigating the relative persuasiveness of different rhetorical techniques found there to be a large degree of heterogeneity in the persuasiveness of different rhetorical elements across policies (Blumenau and Lauderdale, 2022). Therefore, by including a range of policies, I can average over them to address my central questions. Further, prior US-based work has highlighted that voters may make inferences about a politician's party based on the policy in question as the Democrats have "ownership" of particular issues such as education, and the Republicans over others such as defence (Petrocik, Benoit and Hansen, 2003). We therefore may be concerned that voters will evaluate politicians based on their perceptions of parties. In the UK, while certain issues are arguably associated with a particular party – for instance, welfare and the Labour Party (O'Grady, 2022) – I select housing, health, and transport as issues that are central in British politics but not "owned" by either party. Further, the treatments were written to reflect both Conservative and Labour priorities on the issue areas. For instance, the housing treatments include reference to areas that Labour champions, such as the gaps between wages and house prices and protecting vulnerable renters, and issues that the Conservatives have greater authority on, such as increasing homeownership. By doing so, I aim to minimise the likelihood that voters make inferences about a politician's party.

Fourth, the **gender** of the MP: *man* or *woman*. Gender is delivered through fictional Anglo-Saxon names that represent the "typical" politician. Previous experimental work has shown that changing name alone is a sufficient cue to induce voters' gendered attitudes (see Campbell and Cowley, 2014, for a discussion). In total, there are 3 x 2 x 3 x 2 (style x treatment status x policy area x gender) = 36 unique treatment conditions which

serve as the basis of this experimental design. See the appendix for all treatment texts.

To ensure that the speeches are similar in spirit to the kinds of arguments politicians deliver in the UK parliamentary context, the speeches were informed by searching for debates on similar policy areas recorded in Hansard, the official report of parliamentary debates. When writing the speeches, the basic structure remained the same within a policy, however I added words and sentences deemed representative of the style types.[1] For instance, a common approach for politicians delivering anecdotes is by referring to their constituents' experiences (Atkins and Finlayson, 2013). As such, for the anecdote conditions, I refer to the same fictional constituency couple. Table 1 shows an example for each style, where the "treatment" of the style is indicated in bold and square brackets.

To ensure I met these aims, I fielded a pre-test survey through by Prolific to 1,500 members of their UK panel. The purpose was to ensure that the treatments were, and controls were not, considered representative of the styles. Overall, the results are encouraging: 72.5% of respondents assigned to the emotion treatment conditions (strongly) agreed that the treatments were emotional. Similarly, 67.7% of respondents assigned the aggression treatment conditions (strongly) agreed that the treatments were aggressive. While most respondents perceived that both the statistics and anecdote speeches were evidence-based, a larger percentage of respondents perceived the statistical speeches as evidence-based (80.8% and 56.8% respectively). Further, the pre-testing results provide good evidence that the controls were not perceived to be particularly representative of the style types. Only 18.3% of respondents assigned to the emotion control conditions (strongly) agreed that the treatments were emotional. Additionally, 24.4% of the respondents assigned to the aggression control conditions (strongly) agreed that the treatments were aggressive. Full details are in the appendix.

[1]The word choice was informed by the words considered most representative of the styles as measured in recent work on gendered style usage in the UK parliament (Hargrave and Blumenau, 2022).

Table 1: Vignette examples

Attributes	Vignette text
Emotion, treatment, transport, woman	**Lucy Craddock, MP**: *"Transport is the largest carbon-emitting sector of the UK economy, and, within this, cars contribute most. Air pollution increases the risk of heart disease, cancer, diabetes, and asthma attacks. Electric vehicles offer one method of reducing emissions as they produce no air pollution.* ***[Imagine stepping outside on a bright, beautiful morning and hearing not engines revving nor choking on polluted air but feeling that simple joy of hearing the birds sing and that rush of fresh air into your lungs. Doesn't this sound amazing? This could be our future, and I so hope that this doesn't have to be a distant dream.]*** *We should widen accessibility in the use of electric vehicles to make them more practical for those living in urban or built-up areas. If we use electric vehicles, journeys can be greener and safer.* ***[We can fill our lives with the simple pleasures of bird song in our ears, fresh air in our lungs, and blue skies ahead.]"***
Aggression, treatment, health, man	**Jack Richards, MP**: ***["I have to start by saying I utterly disagree with what you've just said. You clearly lack any understanding of how serious this is, and, frankly, you show absolutely no care for the people we represent.]*** *The way we understand diseases and how to treat them has grown over the years. There have been advances in treatments that might help save the lives of people in this country and we should be committed to using them.* ***[But often nothing is done because it is simply deemed not worth the money. I am utterly revolted by the idea that people are not getting treated merely because of some appalling cost-benefit calculation. This is disgusting, deplorable, and inhuman.]*** *It is important that how we fund and pay for drugs works for everybody.* ***[Things must change, and anyone who opposes this cannot claim they care about the well-being of our people."]***
Emotion/aggression, control, housing, woman	**Beth Craddock, MP**: *"Our housing market has not been as it should for years. It is the job of all of us to increase the supply of housing available. But for many young people the gap between wages and house prices is too wide for homeownership to be viable any time soon. Many people live in unstable rented housing who may be driven out by increasing rent costs. Work should be done to help people move out of rented accommodation and become homeowners. We need different strategies to help to increase supply and make homes more affordable. We should build new homes, and we should repurpose empty homes. Builders, investors, and local councils will need to work together for change to occur. Our housing market does not work for many people. There is a need for new policy to help home ownership become realistic for young people all around the country."*

Attributes	Vignette text
Evidence, statistics, housing, man	**Adam Jones, MP**: *"Our housing market has not been as it should for years. It is the job of all of us to increase the supply of housing. But for many young people the gap between wages and house prices is too wide for homeownership to be viable any time soon.* ***[Those in their mid-30s to mid-40s are three times more likely to be renters than 20 years ago. People in the early 1990s could expect to pay 3.5 times their annual earnings on buying a home, but this has risen to 7.8 today. In 2019, the average property sold for £235,300, meanwhile the average pay came in at £29,000.]*** *Our housing market does not work for many people. We need different strategies to help to increase supply and make homes more affordable. There is a need for new policy to help homeownership become realistic for young people all around the country"*
Evidence, anecdote, transport, woman	**Charlotte Richards, MP**: *"Transport is the largest carbon-emitting sector of the UK economy, and, within this, cars contribute most. Air pollution increases the risk of heart disease, cancer, diabetes, and asthma attacks. Electric vehicles offer one method of reducing emissions as they produce no air pollution. The market for electric vehicles is small yet growing. We can be industry leaders in how we produce and use electric vehicles.* ***[I spoke recently to Eleanor and Michael, a young couple in my constituency who live in a flat in a high-rise building. They told me while they want to make the swap to an electric vehicle, it is just not practical as they do not have easy access to a charging point.]*** *We should widen accessibility in the use of electric vehicles to make them more practical for people,* ***[just like Eleanor and Michael]****, who live living in urban or built-up areas. If we use electric vehicles, journeys can be greener and safer."*

I leverage parliamentary speeches as a forum where we might anticipate gender bias. There are, of course, alternative forums, such as politicians' campaign speeches or media descriptions of politicians' behaviour. Alternative channels of communication, such as media reporting, would enable me to assess only how gendered *framing* of politicians may influence voter judgements, and not how voters may form judgements when they engage directly with politicians' speeches. While campaign speeches may seem the more natural forum through which politicians communicate with voters, previous work has highlighted that the single-member district electoral system in the UK provides MPs with an incentive to cultivate a personal vote (Kam, 2009). Further, work on UK parliamentary speech has shown that politicians make strategic use of this forum to appeal to voters (Blumenau, 2021; Osnabrügge, Hobolt and Rodon, 2021). Therefore, in the UK, parliamentary speech is an appropriate forum to study my key questions of interest.

I focus on how voters judge politicians' use of stereotype-(in)congruent styles when they *read* arguments and am therefore unable to make statements about how such dynamics may differ when *watching* politicians. While a central part of style use is captured in the content of speech, tone and body language are important stylistic features. Boussalis et al. (2021) assess whether voters are biased in how they evaluate the styles of Angela Merkel through utilising a variety of video, audio, and text approaches. Their results – that Merkel is punished for aggressiveness and rewarded for happiness – hold across the different style measures but are most pronounced for non-verbal communication. By focusing on written speech, I may understate the extent to which voters are biased in their judgements. Presenting voters with speeches delivered in an audio or video format may, however, introduce unwanted confounders into the relationship – such as difficulties in holding tone, pitch, or cadence constant – which would be challenging to account for. Instead, by focusing on written speech, I can hold constant all features except for politician gender, and identify my key quantity of interest: whether

a politician's gender alone influences voters' perceptions and evaluations.

A further source of concern may be that, although the MP's party is not stated, voters may make inferences about party based on their gender. In the US, voters tend to stereotype the Democrats as more "feminine" and the Republicans as more "masculine" (Winter, 2010), and voters use a candidate's gender to infer candidate ideology (Koch, 2000). US-based experimental work has also emphasised the importance of partisanship in the degree to which voters stereotype politicians (Cassese and Holman, 2018). In the US, therefore, voters have been shown to both infer that women are more liberal than men, and that a candidate's partisanship determines the extent to which they incur negative evaluations. While I am unaware of work that has assessed this question directly in the UK, several factors suggest this is perhaps less of a concern in my context. First, UK right-wing parties have made increasing efforts to integrate women numerically and to better represent women's interests (Childs and Celis, 2018). While Labour made earlier efforts to increase women's representation (Childs, 2004), the Conservatives have also made a concerted effort in recent years to "feminise" the party by incorporating women into the party hierarchy and policy (Childs and Webb, 2012). Second, while Labour have proportionally more women legislators, the Conservatives have had two women Prime Ministers. Therefore, unlike in the US, the parties are more equal with respect to the visibility of women. Finally, recent work has also shown that while there are important differences with respect to party and gender stereotypes in the US, parties in the UK are less divided (Saha and Weeks, 2022). Overall, gender equality is a less polarising and party-political issue in the UK, and the British public has greater familiarity with the presence of women in both major parties. As such, while party is a critical factor with respect to gender and evaluations in the US, this is perhaps less of a concern in the UK political environment.

Survey

I use these treatments as the basis of a vignette survey experiment which was fielded by YouGov to their UK online panel in September 2021. I pre-registered the design, expectations, and analysis plan.[2] The sample was 1,676 people who are nationally representative of the British public on a range of attitudinal and demographic criteria.

Following an introduction screen describing the task, respondents were presented with an argument delivered by a fictional politician. To encourage respondents to read the speech, there was a 15 second delay between the speech presentation and the first question. For a respondent's first task, a style was sampled from the full set of styles (emotion; aggression; evidence). For the selected style, a policy was sampled from the full set of areas (transport; housing; health). Within the policy area, a treatment status was assigned (control; treatment). Finally, the gender of the MP delivering the speech (man; woman) was assigned.

Each respondent completed the task three times. For a given respondent, styles and policies were sample without replacement from round-to-round, such that once a respondent had read a speech representative of a style or policy, they did not read a second speech representative of the same style or policy. For example, if a respondent was assigned emotion on the first response, they could only be assigned aggression or evidence on the second, and the remaining style on their final response. Similarly, if a respondent was assigned transport on the first response, they could only be assigned housing or health on the second, and the remaining policy area on their final response. Style treatment status and gender were sampled with replacement.

[2]The pre-registration documents (Hargrave, 2021) can be found at https://osf.io/bfds7.

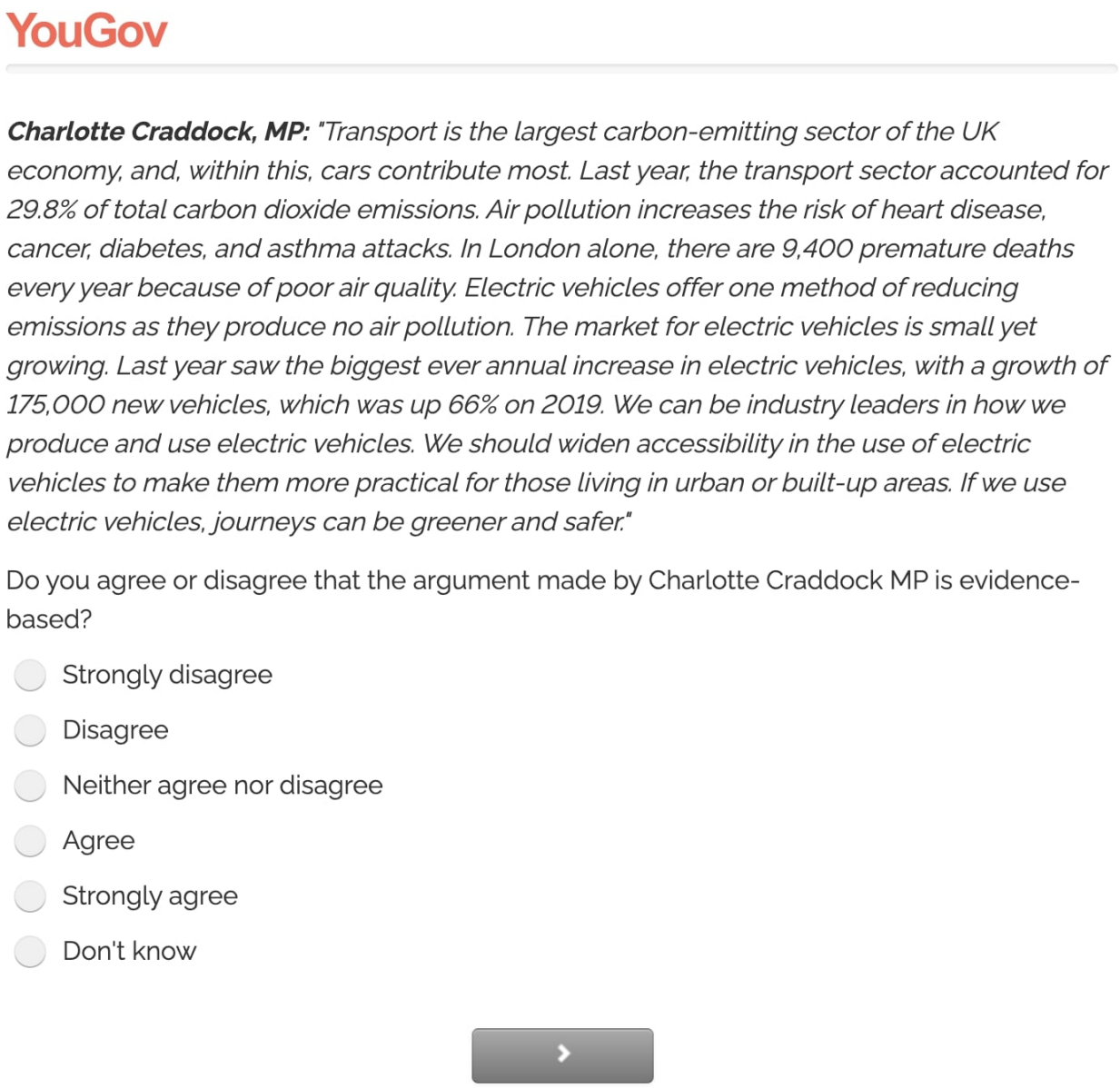

Figure 1: Example experiment prompt (Evidence, statistics, transport, woman)

Per task, each respondent was asked three questions. First, for all styles, respondents were asked whether they "agree or disagree that the MP seems likeable". Second, for all styles, respondents were asked whether they "agree or disagree that the MP seems competent". Third, respondents were asked whether they "agree or disagree that the argument made by the MP is [emotional/aggressive/evidence-based]". Respondents were only asked how emotional, aggressive, or evidence-based they found an argument to

be for the style assigned. For instance, if assigned emotion, they were asked "do you agree or disagree that the argument made by [MP name] is emotional", but not whether they perceived the argument to be aggressive or evidence-based. An example prompt is shown in figure 1.

Given the three observations per respondent and 1,676 respondents, the total number of observations is 5,028 or 1,676 per style.

Methodology

Outcome and explanatory variables

The five outcome variables are each measured on 5-point Likert scales that range from "strongly disagree" to "strongly agree" and include a "don't know" option. Likert scales were selected to maximise respondent interpretation of each of the scale points. First, the **likeability** evaluation of an MP. Second, the **competence** evaluation of an MP. Third, the **perceived emotion** of an argument. Fourth, the **perceived aggression** of an argument. Fifth, the **perceived evidence** of an argument. I drop all "don't know" responses.

There are two main explanatory variables. First, the **gender** of the MP: 0 for a man, 1 for a woman. Second, the **style treatment**, which describes the treatment group status of the style, and takes the value of 0 for the control conditions (non-emotional style, non-aggressive style, and statistical evidence) and 1 for the treatment conditions (emotional style, aggressive style, anecdotal evidence).

Empirical strategy

Responses from the emotion, aggression, and evidence styles are modelled separately as the expected direction of the gender effect differs depending on the style. For each, I estimate a series of OLS regression models to investigate my key quantities of interest. First, to assess whether style usage affects voters' evaluations of politicians and percep-

tions of their arguments, I estimate a series of models for the five outcomes $Y_{i(j)}$ for an individual i in a style j of the following form:

$$Y_{i(j)} = \alpha + \beta_1 TreatmentStyle_j + \epsilon_i \tag{1}$$

where α in each model describes the average emotion, aggression, evidence, likeability, or competence in the control conditions (non-emotional style, non-aggressive style, and statistical evidence), and $\alpha + \beta_1$ describes the same quantities in the treatment conditions (emotional style, aggressive style, and anecdotal evidence).

Second, to assess whether the effects of style usage on voters' evaluations of likeability differ by MP gender, for each style I estimate the following:

$$\begin{aligned} Likeability_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 TreatmentStyle_j + \\ \beta_3 (WomanMP_j \cdot TreatmentStyle_j) + \gamma X_{i(j)} + \epsilon_i \end{aligned} \tag{2}$$

where β_1 in each model describes the difference between men and women when using the control styles (non-emotional, non-aggressive, and statistical evidence). β_2 describes the difference between using control and treatment styles (emotional style, aggressive style, and anecdotal evidence) among men. β_3 is the key quantity of interest, which describes the difference in the effect of using treatment styles compared to control styles for women compared to men. I expect the β_3 coefficients in the emotion and evidence models to be positive, as emotion and anecdotes are female stereotype-congruent styles. In aggression, I expect it to be negative, as aggression is incongruent with feminine stereotypes and women should therefore suffer in likeability evaluations compared to men. X_i is a vector of additional respondent covariates (gender, age, and education).[3]

[3]Across the various models, there are six additional covariates. First, the **policy** area of the argument (categorical: transport; housing; health). Second, **respondent gender** (binary: man; woman). Third,

Third, to assess whether the effects of style usage on voters' evaluations of competence differ by MP gender, for each style I estimate the following:

$$Competence_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 TreatmentStyle_j + \beta_3(WomanMP_j \cdot TreatmentStyle_j) + \gamma X_{i(j)} + \epsilon_i \quad (3)$$

where β_1, β_2 and β_3 describe the same quantities as above. I expect β_3 in the model for emotion to be positive, suggesting that when women conform to feminine stereotypes, they receive a greater competence reward than men. For aggression, I expect β_3 to be negative, suggesting that women incur negative competence evaluations when they violate feminine stereotypes. For evidence, β_3 will be negative as I expect that women are rewarded in competency evaluations when they deliver statistical as opposed to anecdotal arguments. X_i is a vector of additional covariates (gender, age, education, and political attention).

Finally, to assess whether politician gender leads to differences in $PerceivedStyle_{i(j)}$, for each style I estimate the following:

$$PerceivedStyle_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 TreatmentStyle_j + \beta_3(WomanMP_j \cdot TreatmentStyle_j) + \gamma X_{i(j)} + \epsilon_i \quad (4)$$

where β_1, β_2, and β_3 describe the same quantities as above. X_i is a vector of additional covariates (gender, left-right placement, age, and education).

In the empirical strategy described above, I outline numerous statistical tests, and there is risk of the multiple comparisons problem. To assess whether the results are

respondent left-right placement (continuous: 1-7). Fourth, **respondent age** (continuous: age in years). Fifth, **respondent education** (binary: no degree; degree). Sixth, **respondent political attention** (continuous: 0-10). As outlined in the pre-analysis plan, pre-treatment covariates are included as they should, in expectation, increase the precision of the analysis by explaining variation in the outcome variables.

robust, I carry out subsequent analyses with adjusted p-values which control for the False Discovery Rate using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995). The results reported here are unadjusted, and the adjusted p-values are reported in the appendix.

Results

Unconditional effects

In figure 2 I present the results from the models described above in equation 1. There are several findings to note. First, the top panel – which shows the estimated difference between treatment and control styles for the style perception outcomes – highlight that the treatments were successful in shifting perceptions. The aggressive and emotional styles were perceived as significantly more aggressive and emotional than the non-aggressive and non-emotional styles. Further, statistical evidence was perceived as significantly more evidence-based than anecdotal evidence. Voters therefore do perceive the treatments as more representative of the styles than the controls.

Second, the styles that politicians use influences voters' evaluations of their likeability: aggressive politicians are significantly less likeable than non-aggressive politicians, and politicians using anecdotes are significantly more likeable than politicians using statistical evidence. While the point estimate suggests emotional politicians are more likeable than non-emotional politicians, the effect is non-significant.

Third, in the bottom row, I document how style usage influences voters' competence evaluations. Politicians using non-aggressive and non-emotional language are evaluated as more competent than those using aggressive or emotional language. Further, politicians are evaluated as more competent if they use statistical evidence as opposed to anecdotal evidence. Therefore, the styles that politicians use influences how voters evaluate them. However, it seems that certain styles lead to trade-offs in evaluations:

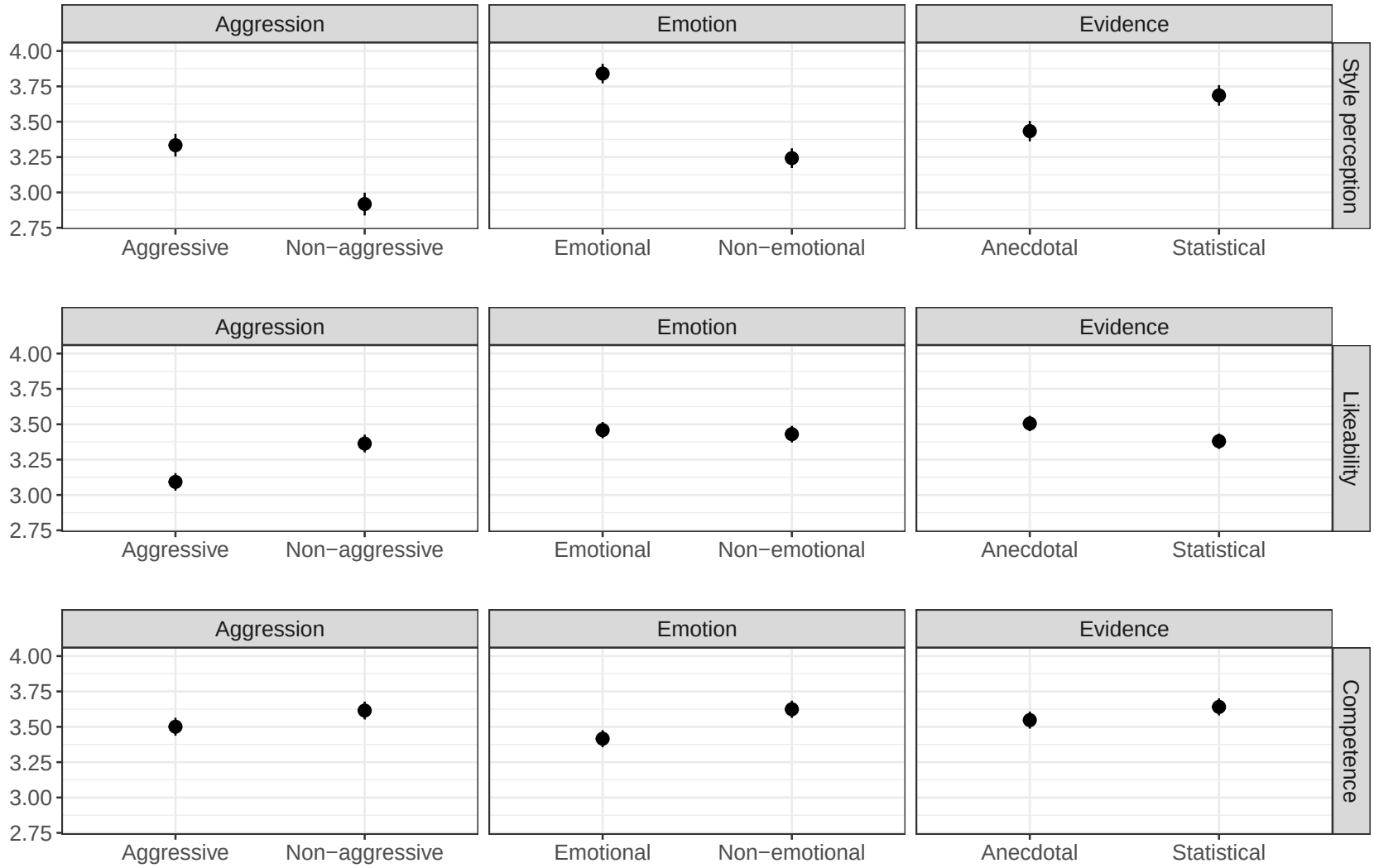

Figure 2: Unconditional relationship between control and treatment style usage, style perceptions, and MP likeability and competence evaluations.

for instance, the use of statistical evidence leads to politicians being perceived as *less* likeable but *more* competent.

Conditional effects by MP gender

Does style usage matter more for women than men? Figure 3 shows the results for the likeability outcomes in the top row. I expected that women would be evaluated as more likeable when they express styles which are congruent with feminine stereotypes and, conversely, suffer in likeability evaluations when they violate stereotypes. The top left panel shows the results for aggression. Here, both men and women politicians are punished for being aggressive. The effect is larger for women than for men, but this difference is non-significant. The top middle panel shows the results for emotion. Here, while the direction of the effect suggests that emotional arguments improve likeability amongst women but not men, the effect again is non-significant. Further, as expected, men are *not* penalised when they use stereotype-incongruent styles (non-emotional). The top right panel shows the results for evidence. I again do not see that women are penalised when they express styles which are incongruent with female stereotypes (statistics) compared to styles which are congruent with female stereotypes (anecdotes).

Consequently, there is no evidence that women politicians are disproportionately penalised in likeability assessments for using styles which are stereotype-incongruent. As figure 2 shows, there is, however, good evidence that the styles politicians use does affect voters' likeability assessments. When politicians use styles consistent with "communal" stereotypes they are perceived as *more* likeable than styles consistent with "agentic" stereotypes.[4] That voters find politicians to be more likeable when they express styles

[4]When estimating a non-pre-registered model pooling all styles together where the outcome is the likeability evaluation and the explanatory variable is whether a style is female stereotype-congruent (emotional style, non-aggressive style, and anecdotal evidence) or not (non-emotional style, aggressive style, and statistical evidence), I find that politicians are perceived as *more* likeable when they express female stereotype-congruent styles.

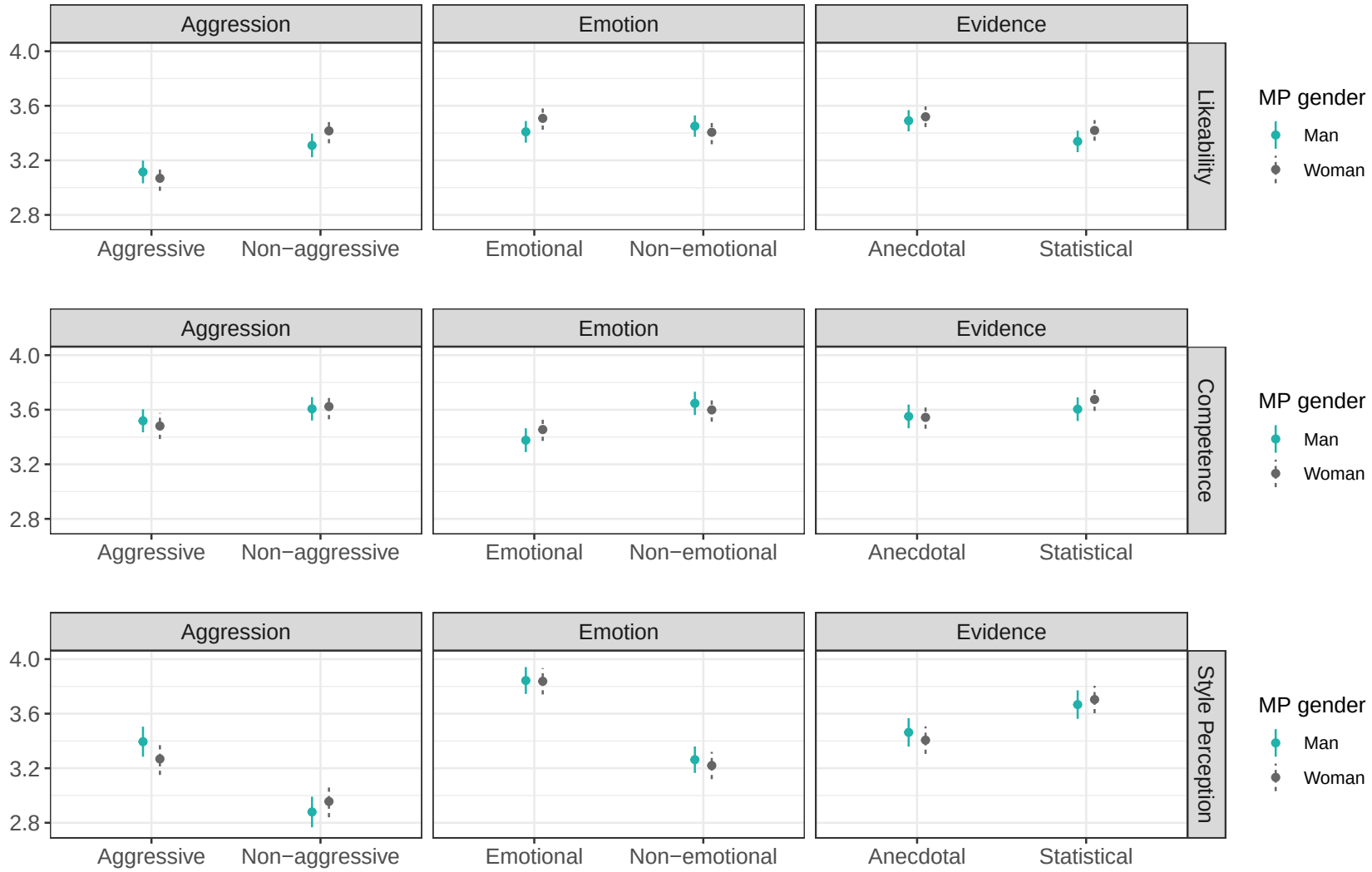

Figure 3: **Conditional relationship between MP gender, style treatment group, style perceptions, and MP likeability and competence evaluations.** The emotional style, non-aggressive style, and anecdotal evidence are female stereotype-congruent, and the non-emotional style, aggressive style, and statistical evidence are female stereotype-incongruent.

consistent with the concept of communality may not be a surprising finding, given that communal stereotypes are associated with being warm, kind, emotional, and people-oriented (Schneider and Bos, 2019).

The middle row of plots in figure 3 show the results for the competence outcomes. In the middle left panel, I show the results for aggression. There is no evidence that either men or women are perceived as more competent when they are aggressive than when they are not. For emotion, there is no evidence that women are evaluated as more competent when they express styles which are congruent with female stereotypes (emotional style). However, the use of female stereotype-congruent styles results in men being evaluated as less competent than female stereotype-incongruent styles. For evidence, I again see that neither men nor women are evaluated as more competent when they use either anecdotal or statistical evidence.

Therefore, I find no evidence that women in particular are penalised in competence evaluations when they express female stereotype-incongruent styles. As with likeability, voters' competency evaluations of politicians are affected by the styles they use. When politicians express styles consistent with "communal" stereotypes they are perceived as *less* competent than styles consistent with "agentic" stereotypes.[5] That voters find politicians to be more competent when they express styles consistent with agentic stereotypes again seems to be intuitive finding given the compatibility between agentic stereotypes and *leadership* stereotypes (Bauer, 2017). The findings for emotion and aggression are consistent with work by Brooks (2011), who finds no double standard in the extent to which voters penalise men and women politicians for their expressions of anger and tears.[6]

[5]When estimating a non-pre-registered model pooling all styles together where the outcome is the competence evaluation and the explanatory variable is whether a style is female stereotype-congruent, I find that politicians are perceived as *less* competent when they express female stereotype-congruent styles.

[6]In the appendix, I also estimate non-pre-registered models pooling styles together where the outcomes are the likeability and competence evaluations, the explanatory variables are a categorical variable

The bottom row in figure 3 shows the results for the style perception models. In the left panel, voters do perceive politicians as more aggressive when they deliver aggressive arguments than non-aggressive arguments, however, there is no evidence that this is gendered. The middle plot shows a very similar result: voters do perceive emotional arguments to be more emotional than non-emotional arguments, but this again is not gendered. Finally, the bottom right plot shows that while voters perceive statistical arguments as more evidence-based than anecdotal arguments, there is again no evidence that this effect is gendered. Voters do not perceive anecdotal or statistical arguments delivered by men to be more evidence-based than equivalent arguments delivered by women. Further, I also expected that differential perceptions of styles might serve as a mechanism through which voters differentially evaluate politicians. I find no evidence of differential effects for style perceptions and, consequently, variation in likeability and competence evaluations are very unlikely to be explained by differential perceptions of styles themselves.

In the theory outlined above, I argued that women would be punished when they communicate in ways which are incongruent with traditional "communal" female stereotypes. However, while I find that the styles politicians use has important consequences for voters' evaluations, I find no evidence that women disproportionately suffer, at least with respect to likeability and competence evaluations, when they express styles which are incongruent with female stereotypes. Further, I find no evidence that voters' percep-

for each style, MP gender, the interaction between the two, and controls for policy areas. This analysis enables me to compare the effect of each of the styles back to the control arguments for each of the policy areas. The upshot of this analysis is consistent with the main results presented here. That is, politicians' style usage influences voters' evaluations of their likeability and competence, but these evaluations are not gendered. I present further details and a full interpretation of this non-pre-registered analysis in the appendix.

tions of the styles themselves are gendered.[7][8]

Conditional effects by voter and MP gender

How might (gendered) judgements of the styles politicians use vary by voter gender? Prior work examining the importance of voter gender on evaluations of politicians has produced inconclusive results. Some work has uncovered no differences between men and women voters (Bauer, 2015*b*), while other work finds that voters may be less likely to penalise politicians from their own gender (Rudman and Goodwin, 2004). To assess heterogeneous effects by voter gender, I carry out two sets of analysis. First, I examine whether men and women voters are equivalently sensitive to politicians' style usage. In the appendix, I assess this by interacting voter gender and style treatment group status for each of the outcomes. Only for the evidence style type are there differences between men and women voters. While men do not find politicians' using anecdotes to be less competent or their arguments to be less evidence-based than statistical evidence, women voters do. Further, while the use of anecdotes improves men voters' likeability assessments relative to the use of statistical evidence, this is not the case for women voters. To the extent that there are differences in how men and women voters evaluate politicians' use of styles, the differences are concentrated among the evidence style.

Second, I assess whether men and women voters are differentially sensitive to the extent to which women politicians conform to stereotype-congruent behaviours. In the

[7] Although not the main quantity of interest, it is possible that men and women may receive differential evaluations in likeability and competence depending on the policy in question. In the appendix, I estimate a non-pre-registered analysis to assess whether this is the case, and find no evidence of differential evaluations for policy areas.

[8] My primary quantities of interest are interaction effects between style treatment status and MP gender that are non-significant. A plausible concern is that the design I present is insufficiently powered to detect the effect sizes reported. To address this, in the appendix, I report the results of a power analysis where I simulate the data collection process for the fixed sample size available for different hypothetical standardised effect sizes ranging from very small to large effects according to conventional standards. I find that if the true effect size was small then my design may not be sufficiently powered to detect this effect, however the sample size I have is sufficient to detect medium effect sizes.

appendix, I subset the data into men and women voters and replicate the main analysis described above. While for aggression and evidence there are no differences, this is not the case for emotion. For likeability, women politicians in particular are rewarded for expressing emotional styles instead of non-emotional styles among women voters; an effect I do not find for men voters. For competence, I find that while women voters find emotional politicians overall to be less competent than non-emotional, they give women *more* of a competency reward than men for expressing emotional styles. I again see no such effect among men voters. Finally, for perceived styles, I see that women politicians are perceived as particularly emotional compared to men politicians when they express emotional styles; an effect I do not see for men respondents.

Therefore, women voters give a larger likeability and competence reward to women politicians who are emotional and perceive women politicians as more emotional than men politicians. I find no evidence of these effects for men voters. Consequently, to the extent that there is any evidence of women politicians being rewarded for conforming to stereotype-congruent styles, the effects are concentrated amongst women voters and emotion.

Conclusion

Do the styles politicians use influence how voters evaluate them, and does this matter more for women than for men? In this paper, I address these questions through a novel survey experiment where I present UK voters with speeches where the argument style and gender of the politician delivering the argument are varied. This enables me to identify, first, whether politicians experience a backlash effect with respect to evaluations of likeability and competence when they deploy styles that are gender stereotype-incongruent, and second, whether voters' differential perceptions of the styles themselves might explain this backlash.

I report four main findings. First, style usage has important consequences for voters' evaluations of politicians. Politicians who are unaggressive and draw on anecdotes are more likeable, whereas politicians who are unemotional and reference statistics are more competent. Second, I find no evidence that voter evaluations of politicians are gendered. In particular, women politicians are *not* punished for stereotype-incongruent behaviour. Third, while there is clear evidence that voters can identify the styles politicians use, I also find no evidence that voters' perceptions of the styles themselves are gendered. Gender bias in voters' perceptions of the styles themselves is very unlikely to explain variation in the likeability and competence evaluations. Fourth, I find some evidence that these evaluations differ by voter gender.

Across the various styles and outcomes, the main finding I document is therefore that styles influence voters' evaluations of politicians, but these evaluations do *not* vary by MP gender. Why do I find little evidence of gender bias in voters' evaluations, and what implications may these findings have for voting behaviour? First, while I find that style usage influences voters' evaluations of the likeability and competence of politicians, these findings do not, however, enable me to assess whether these evaluations have downstream consequences for voting behaviour. The styles and personalities of leaders have long been considered an important determinant of voters' attitudes (Declercq, Hurley and Luttbeg, 1972). Further, as partisanship in the electorate has declined over time (Dalton and Wattenberg, 2000), and voters have become increasingly volatile (Fieldhouse et al., 2019), then the styles politicians express and associated evaluations may have even increased in importance as determinants of vote choice as voters base their decisions on factors beyond party. Further, while I cannot directly assess whether these evaluations influence vote decisions, previous UK-based work has shown that voters' evaluations of politicians' competency do influence their voting preferences (Green and Jennings, 2017). It is therefore not unreasonable to assume that likeability and com-

petence evaluations may inform vote intention.

My findings suggest that politicians may face trade-offs in evaluations: while styles compatible with communality lead to positive likeability evaluations, styles compatible with agency lead to positive competency evaluations. Should politicians prioritise competence or likeability evaluations? Traditional accounts of leadership have suggested that competence is important in informing vote choice (Green and Jennings, 2017), and congruent with the traits and behaviours deemed necessary and suitable for leaders (Elgie, 2015). As such, according to traditional accounts, we may consider competence the more important evaluation to optimise. However, a trend that is common across many political contexts in recent decades is that voters are increasingly dissatisfied with politics, and find that politicians are out-of-touch and unlike "normal people" (Clarke et al., 2018). In response to voter dissatisfaction with this type of politics and politician, there is an increasing desire for politicians who are instead human, personable, charismatic, engaging, and in-touch (Valgarðsson et al., 2021). The move to telegenic, human, and personable styles has been said to be a core element of the success strategies of populist candidates (De Vries and Holbolt, 2020), and recent examples of the use of these styles, such as the widespread fame and popularity of Ukrainian President Volodymyr Zelensky, suggests that voters and the media find this to be compelling.[9] As such, while traditional accounts may suggest that ensuring competency is more important in determining candidate success, voters have begun to place increasing importance on the likeability of politicians.

As far as I am aware, we currently lack systematic evidence on which evaluations of politicians among the many that prior work has studied, such as likeability, competence, honesty, hard-workingness, or charisma, matter most in informing voting behaviour. A fruitful avenue for future work would, therefore, be to identify which traits voters priori-

[9]See, for example, "The Zelensky Effect: How To Engage, Energize and Unleas Your Organization's Potential", *Forbes*, 29th March 2022.

tise in their decision at the ballot box. Such a study would be useful not only for understanding the wider implications of the results I present here, but also for the plethora of experimental studies that assess how a variety of features of politicians' behaviour – such as whether they are corrupt (Eggers, Vivyan and Wagner, 2017) or loyal to their party (Campbell et al., 2019) – influences how voters evaluate them.

Finally, at the core of the idea that women politicians face double standards when they violate stereotypically expected behaviours is that voters actually hold these expectations for women's behaviour in the first place. However, studies show that the public's perception of the validity of these stereotypes has shifted over time, as women have been seen as increasingly agentic (Eagly et al., 2020). Voters in the UK have also become more gender-egalitarian in their attitudes (Taylor and Scott, 2018), and have become markedly less likely to support traditional gendered divisions in social roles (Shorrocks, 2018). Further, there is evidence that UK politicians have come to behave in a way that is less consistent with traditional gender stereotypes. Women politicians have decreasingly used "communal" and increasingly use "agentic" styles over time (Hargrave and Blumenau, 2022). The pessimistic assumption is that this behaviour change might be met by backlash from voters. However, the results presented here suggest that this may not materialise, as UK voters do not seem to unjustly penalise women for stereotype-incongruent behaviour.

Of course, without a study from 20 years ago to compare these findings to, it is not possible to know whether UK voters in previous eras did apply these descriptive stereotypes or punish women politicians for stereotype-incongruent behaviour. Yet, if voters no longer hold the same stereotypical expectations about men's and women's behaviour, and politicians decreasingly behave in accordance with traditional stereotypes, it may not be surprising to uncover that women are not punished for behaviour that violates traditional stereotypes.

References

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.

Bateson, Regina. 2020. "Strategic Discrimination." *Perspectives on Politics* 18(4):1068–1087.

Bauer, Nichole M. 2015*a*. "Emotional, Sensitive, and Unfit for Office? Gender Stereotype Activation and Support Female Candidates." *Political Psychology* 36(6):691–708.

Bauer, Nichole M. 2015*b*. "Who Stereotypes Female Candidates? Identifying Individual Differences in Feminine Stereotype Reliance." *Politics, Groups, and Identities* 3(1):94–110.

Bauer, Nichole M. 2017. "The Effects of Counterstereotypic Gender Strategies on Candidate Evaluations." *Political Psychology* 38(2):279–295.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society* 57(1):289–300.

Blumenau, Jack. 2021. "Online Activism and Dyadic Representation: Evidence from the UK E-Petition System." *Legislative Studies Quarterly* 46(4):889–920.

Blumenau, Jack and Benjamin E. Lauderdale. 2022. "The Variable Persuasiveness of Political Rhetoric." *American Journal of Political Science* (Forthcoming):1–16.

Boussalis, Constantine, Travis G. Coan, Mirya R. Holman and Stefan Müller. 2021. "Gender, Candidate Emotional Expression, and Voter Reactions During Televised Debates." *American Political Science Review* 115(4):1242–1257.

Brooks, Deborah Jordan. 2011. "Testing the Double Standard for Candidate Emotionality: Voter Reactions to the Tears and Anger of Male and Female Politicians." *The Journal of Politics* 73(2):597–615.

Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes Do Not Harm Women Candidates*. New Jersey: Princeton University Press.

Campbell, Rosie and Philip Cowley. 2014. "What Voters Want: Reactions to Candidate Characteristics in a Survey Experiment." *Political Studies* 62(4):745–765.

Campbell, Rosie, Philip Cowley, Nick Vivyan and Markus Wagner. 2019. "Legislator Dissent as a Aalence Signal." *British Journal of Political Science* 49(1):105–128.

Cassese, Erin C. and Mirya R. Holman. 2018. "Party and Gender Stereotypes in Campaign Attacks." *Political Behavior* 40:785–807.

Childs, Sarah. 2004. *New Labour's Women MPs: Women Representing Women*. London: Routledge.

Childs, Sarah and Karen Celis. 2018. "Conservatism and Women's Political Representation." *Politics & Gender* 14(1):5–26.

Childs, Sarah and Paul Webb. 2012. *Sex, Gender and the Conservative Party: From Iron Lady to Kitten Heals*. Basingstoke: Palgrave Macmillan.

Clarke, Nick, Will Jennings, Jonathan Moss and Gerry Stoker. 2018. *The Good Politician: Folk Theories, Political Interaction, and the Rise of Anti-Politics*. Cambridge: Cambridge University Press.

Dalton, Russell J. and Martin P. Wattenberg. 2000. Unthinkable Democracy: Political Change in Advanced Industrial Democracies. In *Parties without Partisans*. Oxford: Oxford University Press pp. 3–19.

de Geus, Roosmarijn A., John R. McAndrews, Peter John Loewen and Aaron Martin. 2021. "Do Voters Judge the Performance of Female and Male Politicians Differently? Experimental Evidence from the United States and Australia." *Political Research Quarterly* 74(2):302–316.

De Vries, Catherine E. and Sara B. Holbolt. 2020. *Political Entrepreneurs: The Rise of Challenger Parties in Europe*. New Jersey: Princeton University Press.

Declercq, Eugene, Thomas L. Hurley and Norman R. Luttbeg. 1972. "Voting in American Presidential Elections, 1956-1972." *American Politics Quarterly* 3:222–243.

Dietrich, Bryce J., Matthew Hayes and Diana Z. O'Brien. 2019. "Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech." *American Political Science Review* 113(4):941–962.

Ditonto, Tessa. 2017. "A High Bar or a Double Standard? Gender, Competence, and Information in Political Campaigns." *Political Behavior* 39:301–325.

Donnelly, Kristin, Jean M. Twenge, Malissa A. Clark, Samia K. Shaikh, Angela Beiler-May and Nathan T. Carter. 2016. "Attitudes Toward Women's Work and Family Roles in the United States, 1976–2013." *Psychology of Women Quarterly* 40(1):41–54.

Eagly, Alice H, Christa Nater, David I Miller, Michèle Kaufmann and Sabine Sczesny. 2020. "Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of U.S. Public Opinion Polls From 1946 to 2018." *American Psychologist* 75(3):301–315.

Eagly, Alice H. and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice Toward Female Leaders." *Psychological Review* 109(3):573–598.

Eagly, Alice H. and Wendy Wood. 2012. "Biosocial Construction of Sex Differences and Similarities in Behavior." *Advances in Experimental Social Psychology* 46:55–123.

Eggers, Andrew C., Nick Vivyan and Markus Wagner. 2017. "Corruption, Accountability, and Gender: Do Female Politicians Face Higher Standards in Public Life?" *Journal of Politics* 80(1):321–326.

Elgie, Robert. 2015. *Studying Political Leadership: Foundations and Contending Accounts*. Basingstoke: Palgrave Macmillan.

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Jonathan Mellon, Christopher Prosser, Hermann Schmitt and Cees van der Eikj. 2019. *Electoral Shocks: The Volatile Voter in a Turbulent World*. Oxford: Oxford University Press.

Fox, Richard L. and Zoe M. Oxley. 2003. "Gender Stereotyping State Executive Elections: Candidate Selection and Success." *Journal of Politics* 65(3):833–50.

Gill, Michael J. 2004. "When Information Does Not Deter Stereotyping: Prescriptive Stereotyping Can Foster Bias Under Conditions That Deter Descriptive Stereotyping." *Journal of Experimental Social Psychology* 40(5):619–632.

Gleason, Shane A. 2020. "Beyond Mere Presence: Gender Norms in Oral Arguments at the U.S. Supreme Court." *Political Research Quarterly* 73(3):596–608.

Green, Jane and Will Jennings. 2017. *The Politics of Competence: Parties, Public Opinion and Voters*. Cambridge: Cambridge University Press.

Grey, Sandra. 2002. "Does Size Matter? Critical Mass and New Zealand's Women MPs." *Parliamentary Affairs* 55:19–29.

Hargrave, Lotte. 2021. "A Double Standard? Gender Bias in Voters' Perceptions of Political Arguments." *EGAP* (ID: 20210908AB).
URL: *osf.io/bfds7*

Hargrave, Lotte and Jack Blumenau. 2022. "No Longer Conforming to Stereotypes? Gender, Political Style, and Parliamentary Debate in the UK." *British Journal of Political Science* (Forthcoming):1–18.

Hargrave, Lotte and Tone Langengen. 2021. "The Gendered Debate: Do Men and Women Communicate Differently in the House of Commons?" *Politics & Gender* 17(4):580–606.

Hornikx, Jos. 2018. "Combining Anecdotal and Statistical Evidence in Real-Life Discourse: Comprehension and Persuasiveness." *Discourse Processes* 55(3):324–336.

Huddy, Leonie and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and Female Candidates." *American Journal of Political Science* 37(1):119–147.

Jamieson, Kathleen. 1995. *Beyond the Double Bind: Women and Leadership*. Oxford: Oxford University Press.

Kam, Christopher J. 2009. *Party Discipline and Parliamentary Politics*. Cambridge: Cambridge University Press.

Koch, Jeffrey W. 2000. "Do Citizens Apply Gender Stereotypes to Infer Candidates' Ideological Orientations?" *The Journal of Politics* 62(2):414–429.

Mattei, Laura R. Winsky. 1998. "Gender and Power in American Legislative Discourse." *The Journal of Politics* 60(2):440–461.

O'Grady, Tom. 2022. *The Transformation of British Welfare Policy: Politics, Discourse, and Public Opinion*. Oxford: Oxford University Press.

Okimoto, Tyler G. and Victoria L. Brescoll. 2010. "The Price of Power: Power Seeking and Backlash Against Female Politicians." *Personality and Social Psychology Bulletin* 36(7):923–936.

Osnabrügge, Moritz, Sara B. Hobolt and Toni Rodon. 2021. "Playing to the Gallery: Emotive Rhetoric in Parliaments." *American Political Science Review* 115(3):885–899.

Petrocik, John R., William L. Benoit and Glenn J. Hansen. 2003. "Issue Ownership and Presidential Campaigning , 1952-2000." *Political Science Quarterly* 118(4):599–626.

Rudman, Laurie A. and Stephanie A. Goodwin. 2004. "Gender Differences in Automatic In-Group Bias: Why Do Women Like Women More Than Men Like Men?" *Journal of Personality and Social Psychology* 87(4):494–509.

Saha, Sparsha and Ana Catalano Weeks. 2022. "Ambitious Women: Gender and Voter Perceptions of Candidate Ambition." *Political Behavior* 44:779–805.

Schneider, Monica C. and Angela L. Bos. 2019. "The Application of Social Role Theory to the Study of Gender in Politics." *Political Psychology* 40(1):173–213.

Schwarz, Susanne and Alexander Coppock. 2022. "What Have We Learned about Gender from Candidate Choice Experiments? A Meta-Analysis of Sixty-Seven Factorial Survey Experiments." *Journal of Politics* 84(2).

Shorrocks, Rosalind. 2018. "A Feminist Generation? Cohort Change in Gender-Role Attitudes and the Second-Wave Feminist Movement." *International Journal of Public Opinion Research* 30(1):125–145.

Taylor, Eleanor Attar and Jacqueline Scott. 2018. "Gender: New consensus or continuing battleground?" *British Social Attitudes 35* pp. 56–85.

Valgarðsson, Viktor Orri, Nick Clarke, Will Jennings and Gerry Stoker. 2021. "The Good Politician and Political Trust: An Authenticity Gap in British Politics?" *Political Studies* 69(4):858–880.

Winter, Nicholas J.G. 2010. "Masculine Republicans and Feminine Democrats: Gender and Americans' Explicit and Implicit Images of the Political Parties." *Political Behavior* 32(4):587–618.