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

Contents

Experiment pre-test survey **S2**

Treatment texts **S8**

Full models **S15**

Unconditional effects S15

Conditional effects by MP gender S17

Multiple comparisons correction **S23**

Addditional non-pre-registered analyses **S26**

Pooled models S26

MP gender and policy area S32

Heterogeneous effects by voter gender **S35**

Conditional relationships by voter gender S35

Conditional relationships by voter and MP gender S39

Power analysis **S43**

Experiment pre-test survey

Prior to fielding the survey experiment, a pre-test survey was fielded by Prolific to 1,499 members of the UK online panel in August 2021. Speeches were pre-tested without the MP's gender stated, and there were 3 x 2 x 2 (style x style treatment status x policy area) = 18 speeches tested overall.

Following an introduction screen describing the task, respondents were randomly assigned to read one speech and asked to answer several questions about the speech and the MP. First, depending on the style type assigned, respondents were asked whether they agreed or disagreed that the speech was "emotional", "aggressive" or "evidence-based". The purpose of this was to ensure that the treatment texts were representative of the style types, and that the control texts were not representative of the style types. An example of the task and question can be seen in figure S1.

Results from the style pre-testing exercise are presented in table S1. Responses are presented as the extent to which respondents (strongly) disagree, neither agree nor disagree,or (strongly) agree that the speech was "emotional", "aggressive" or "evidence-based". Figures S2-S4 show the responses collapsed by style, style prevalence, and policy area.

Table S1: Results from treatment pre-testing

Style type	(Strongly) disagree	Neither agree nor disagree	(Strongly) agree
Emotion control	60%	21.7%	18.3%
Emotion treatment	10.1%	17.4%	72.5%
Aggression control	65.2%	10.4%	24.4%
Aggression treatment	19.9%	12.4%	67.7%
Evidence statistics	6.4%	12.8%	80.8%
Evidence anecdote	28.8%	14.4%	56.8%

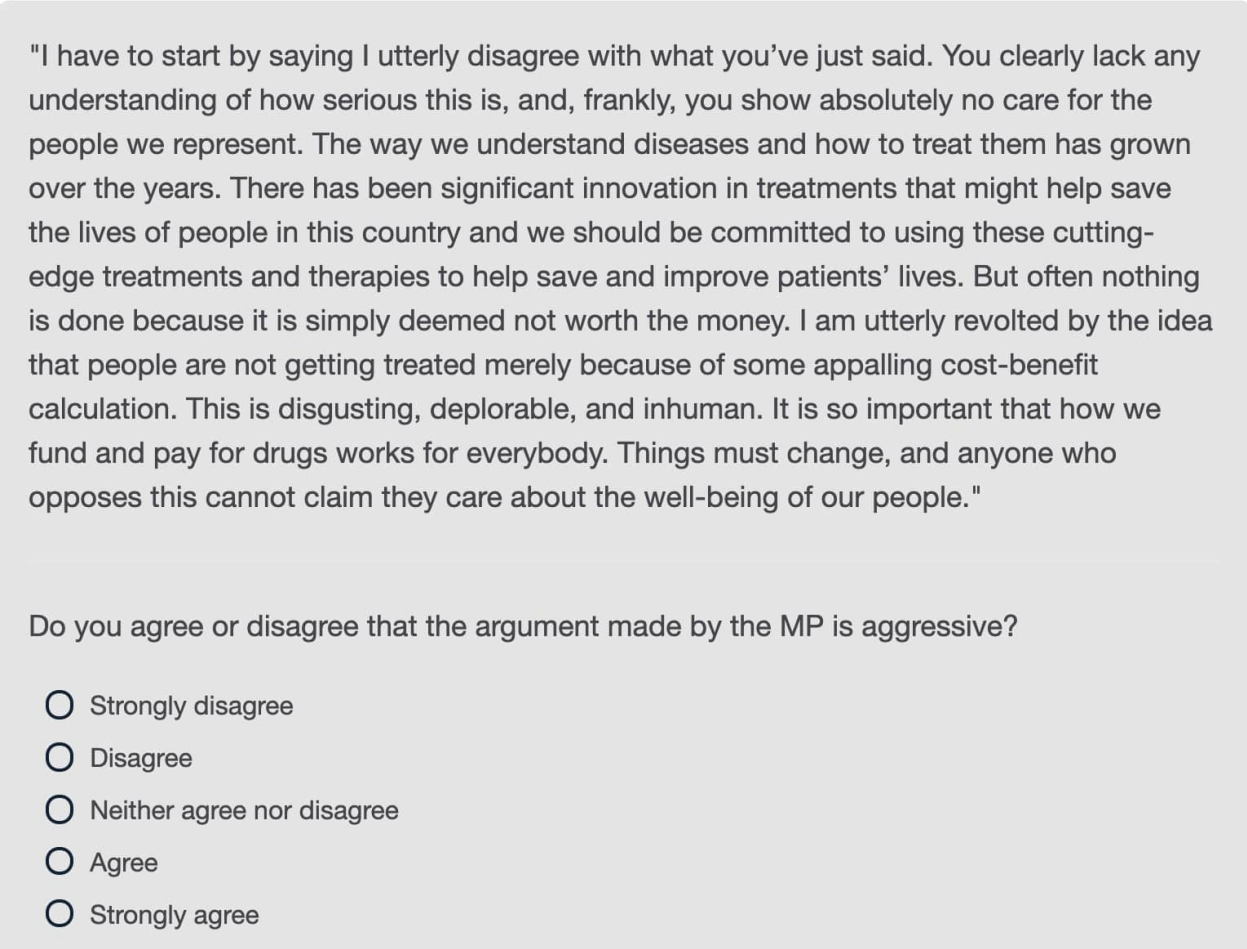

Figure S1: Example of pre-testing task (Aggression, treatment, health)

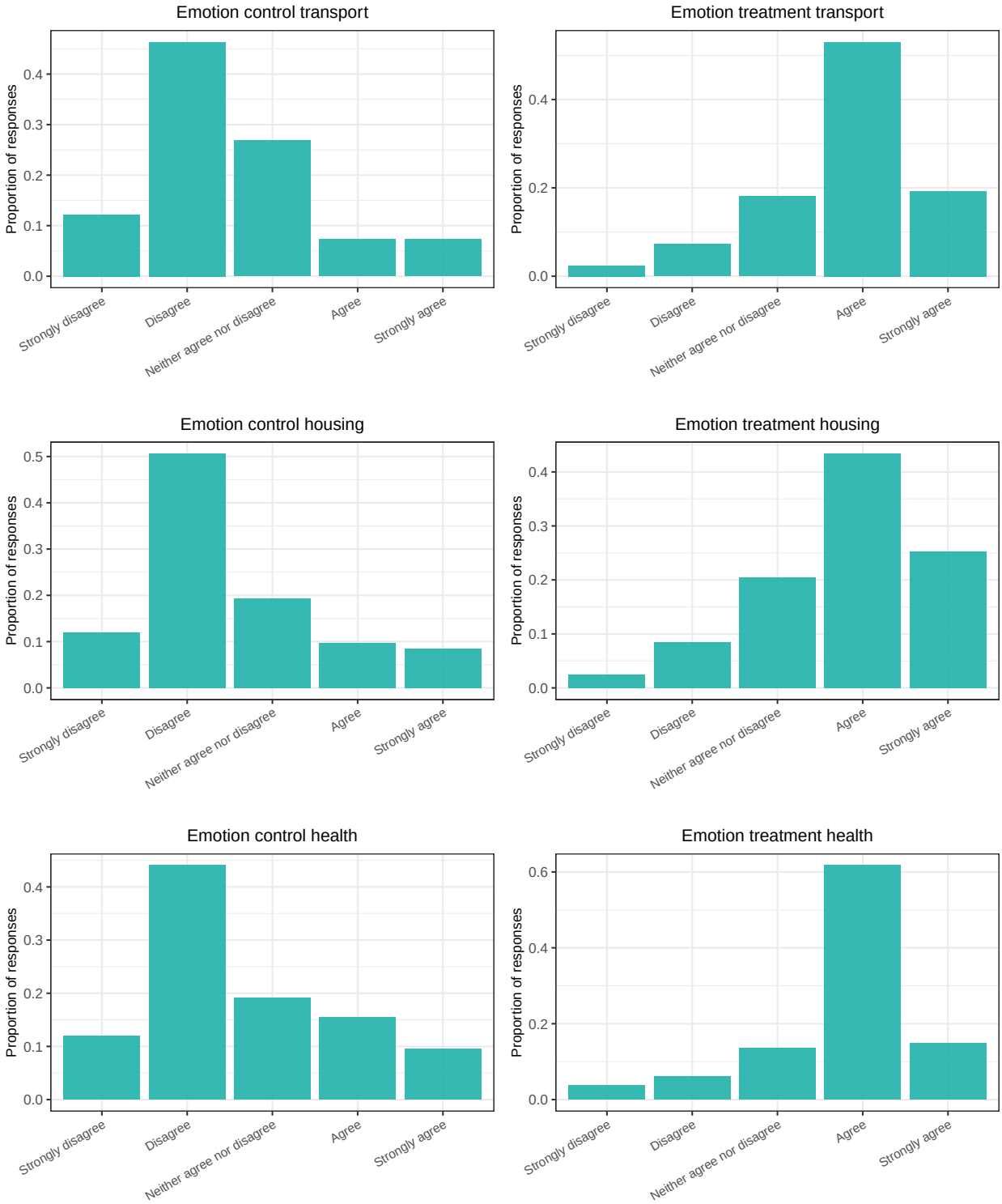

Figure S2: Distribution of responses across treatment group status for emotion

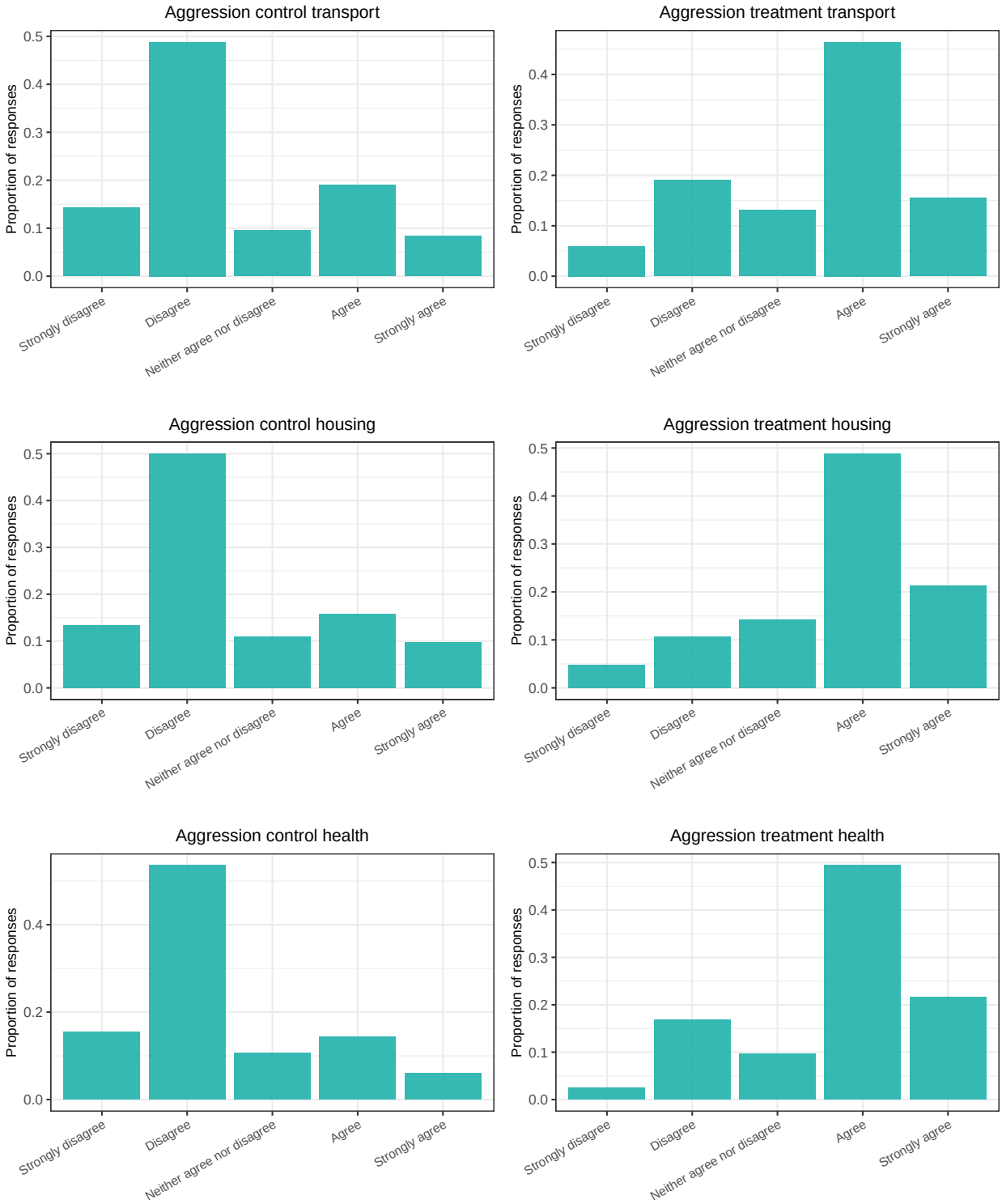

Figure S3: Distribution of responses across treatment group status for aggression

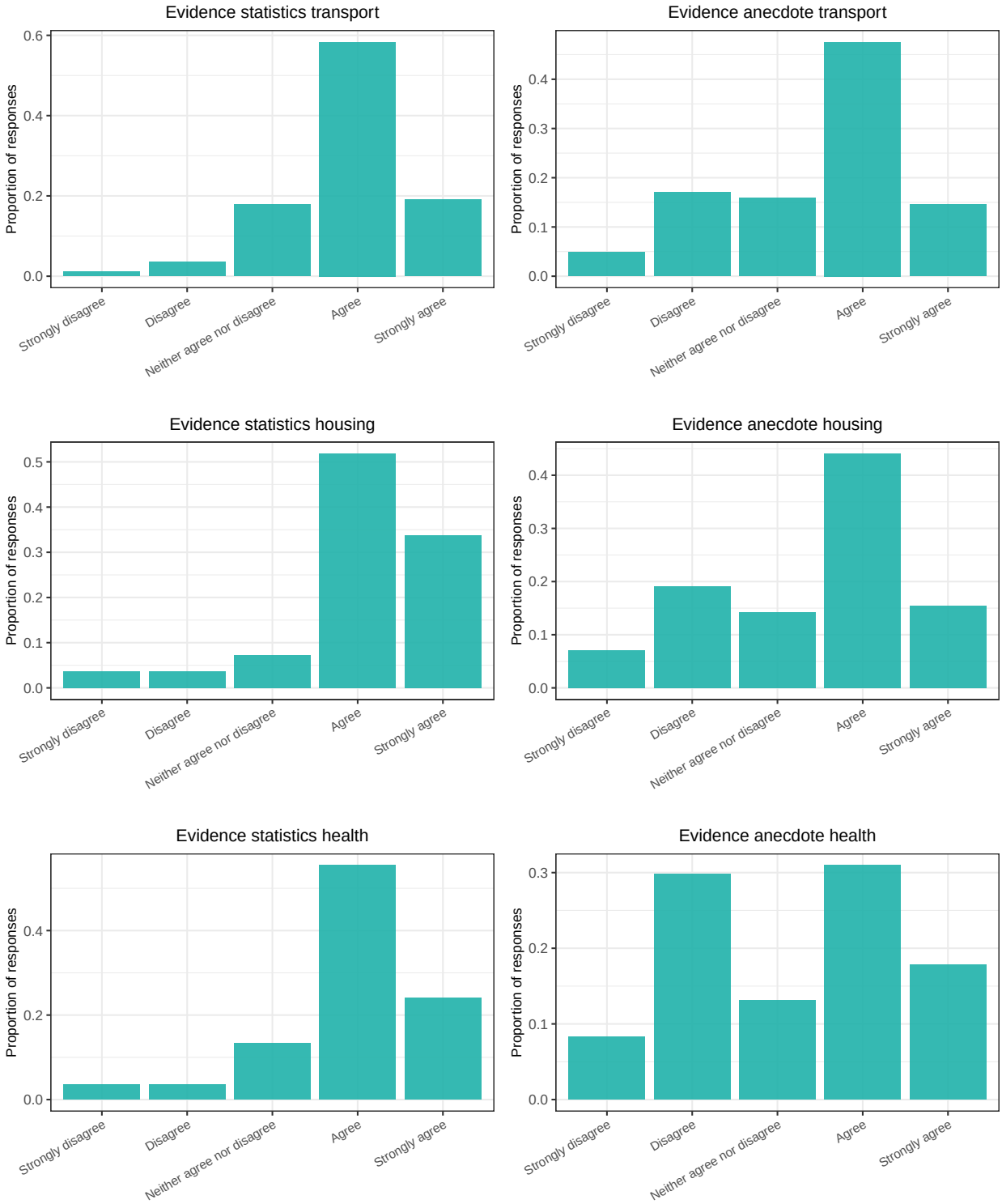

Figure S4: Distribution of responses across treatment group status for evidence

Overall, the results from the style pre-testing exercise confirm that the treatments are perceived as representative of the styles. For instance, 72.5% of respondents assigned to the emotion treatment conditions either agreed or strongly agreed that they were emotional. Similarly, 67.7% of respondents assigned the aggression treatment conditions agree or strongly agreed that they were aggressive. Further, the pre-testing results provide good evidence that the control texts were not perceived to be particularly representative of the styles. For instance, 60% of respondents assigned to the emotion control conditions either disagreed or strongly disagreed that they were emotional. Additionally, 65.2% of the respondents assigned to the aggression control conditions either disagreed or strongly disagreed that they were aggressive. While the majority of respondents perceived that both the statistics and anecdote speeches were evidence-based, a larger percentage of respondents perceived the statistical speeches as evidence based than the anecdotal speeches (80.8% and 56.8% respectively).

Further, the results are consistent across the policy areas for each style. For example, for the emotion controls, between 45%–50% of respondents "disagreed" that the arguments were emotional across the three policy areas, and roughly 10% of respondents "strongly disagreed" that the arguments were emotional across the three policy areas. Similarly, for the aggression treatments between roughly 45%–50% of respondents "agreed" that the arguments were aggressive.

Second, to ensure that treatment texts were not too long for respondents to read and understand, all respondents were asked the follow-up question "Did you find this text too long and/or too complicated to understand?". Overall, across all 18 treatments, 95.7% said that the speeches were not too long and/or complicated to understand, and only 4.3% said they were. Overall, there is good evidence that the style treatments work as planned, and that the treatment texts are not too long for respondents to read and understand.

Treatment texts

In full set of gendered names used in the experiment are displayed in table S2.

Table S2: Gendered names for experimental design

Man forename	Woman forename	Surname
Adam	Beth	Craddock
Jack	Charlotte	Jones
Peter	Lucy	Richards

In table S3 I include a definition of each of the styles of interest, which are drawn from the gender stereotypes literature.

Table S3: Style type definitions

Style	Definition
Emotion	Positive emotional language, which includes expressing empathy, praise, celebration, congratulations, hope, and joy
Aggression	Language that relates to conflict, political point-scoring, criticisms, or insults
Statistical evidence	Use of numbers, statistics, numeric quantifiers, figures, and empirical evidence as the basis for an MP's argument
Anecdotal evidence	Use of personal examples or experiences, stories of other people, constituency stories, or illustrative examples as the basis of an MP's argument

In table S4, I present the full text of each of the speeches. The "treatment" for each of the styles are indicated in square brackets and in bold font.

Table S4: Treatment texts

Treatment number	Treatment text
Treatment 1: Emotion, Control, Housing	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.
Treatment 2: Emotion, Treatment, Housing	**[The idea of owning your own home is one filled with such excitement for people.**] It is the job of all of us to increase the supply of housing. **[I recall with such warmth that sense of euphoria, exhilaration, and real achievement when I held the keys to my first home. I can, hand on my heart, say that it was one of the proudest days of my life.]** But for many young people the gap between wages and house prices is too wide for homeownership to be viable any time soon. We need different strategies to help to increase supply and make homes more affordable. **[I am hopeful that, if we work together, we can help people get closer to experiencing the joy of homeownership themselves. Changing the lives of the people I represent to help reach brighter futures is, after all, really what makes my job such a pleasure to do.]**
Treatment 3: Aggression, Control, Housing	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.

Treatment number	Treatment text
Treatment 4: Aggression, Treatment, Housing	**[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.]** 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. **[Too many people live in insecure rented houses owned by exploitative, greedy, penny-pinching landlords who milk their tenants for all they're worth.]** We need different strategies to help to increase supply and make homes more affordable. **[Young people have been utterly abandoned and left fearful about their futures because we have catastrophically failed to do enough to help them. Things must change, and anyone who opposes this cannot claim they care about the well-being of our people.]**
Treatment 5: Evidence, Statistics, Housing	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.
Treatment 6: Evidence, Anecdote, Housing	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. **[I spoke recently to Eleanor and Michael, a young couple in my constituency, who have been renting together for a few years and want to buy their first home. They told me they are extremely concerned that their dream of homeownership simply will not be one they can reach unless something changes.]** 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, **[just like Eleanor and Michael]**, all around the country.

Treatment number	Treatment text
Treatment 7: Emotion, Control, Health	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 in some cases, patients are left without access to the drugs they need because of the costs of treatment. Patients being able to access the treatments they require is one of the principles of providing universal health care to everyone who needs it. I think some people find it challenging to accept that questions of money should enter decisions about health care, but this is the situation that many hospitals, doctors, and patients find themselves in. It is important that how we fund and pay for drugs works for everybody. Change should happen to make sure that both medical services and patients get the treatments and resources they need.
Treatment 8: Emotion, Treatment, Health	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. **[These advances are miraculous. I am brought to tears when I hear heartening stories about young children's lives being saved and getting to see the joy and relief of families getting to spend more time with their loved ones. Being with the ones we love is really the only thing that matters.]** But in some cases, patients are left without access to the drugs they need because of the costs of treatment. It is important that how we fund and pay for drugs works for everybody. **[This isn't only about improving our country for the better, but about ensuring that we can be hopeful about a future of humanity that shines bright.]**
Treatment 9: Aggression, Control, Health	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 in some cases, patients are left without access to the drugs they need because of the costs of treatment. Patients being able to access the treatments they require is one of the principles of providing universal health care to everyone who needs it. I think some people find it challenging to accept that questions of money should enter decisions about health care, but this is the situation that many hospitals, doctors, and patients find themselves in. It is important that how we fund and pay for drugs works for everybody. Change should happen to make sure that both medical services and patients get the treatments and resources they need.

Treatment number	Treatment text
Treatment 10: Aggression, Treatment, Health	**[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.]**
Treatment 11: Evidence, Statistics, Health	The way we understand diseases and how to treat them has grown over the years. But in some cases, patients are left without access to the drugs they need because of the costs of treatment. I think some people find it challenging to accept that questions of money should enter decisions about health care, but this is the situation that many hospitals, doctors, and patients find themselves in. **[The National Health Service has a fixed budget of only around £110 billion a year in England.]** It is important that how we fund and pay for drugs works for everybody. **[For example, between 2008 and 2016, one drug increased in price by 12,000%. This sort of increase isn't sustainable, and if the price had only stayed the same, the NHS could have spent £58 million less.]** Change should happen to make sure that both medical services and patients get the treatments and resources they need.
Treatment 12: Evidence, Anecdote, Health	The way we understand diseases and how to treat them has grown over the years. **[I spoke recently to Eleanor and Michael, a young couple in my constituency, and they told me the story of how Michael was able to get the treatment he needed to be declared cancer free and for his life to be saved.]** But in other cases, patients are left without access to the drugs they need because of the costs of treatment. I think some people find it challenging to accept that questions of money should enter decisions about health care, but this is the situation that many hospitals, doctors, and patients find themselves in. It is important that how we fund and pay for drugs works for everybody, **[just like it did for Eleanor and Michael]**. Change should happen to make sure that both medical services and patients get the treatments and resources they need.

Treatment number	Treatment text
Treatment 13: Emotion, Control, Transport	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. We should aim for all new vehicles made to be run either partially or wholly on electricity. Electric vehicles offer clear benefits for improving local air quality as they produce no exhaust emissions at the street level. The market for electric vehicles is small yet growing. We can be industry leaders in how we produce and use electric vehicles. A variety of different strategies should be employed to encourage their uptake. We should widen accessibility in the use of electric vehicles to make them more practical for those living in urban or built-up areas. For instance, by making sure there is sufficient charging infrastructure. If we use electric vehicles, journeys can be greener and safer.
Treatment 14: Emotion, Treatment, Transport	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.]**
Treatment 15: Aggression, Control, Transport	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. We should aim for all new vehicles made to be run either partially or wholly on electricity. Electric vehicles offer clear benefits for improving local air quality as they produce no exhaust emissions at the street level. The market for electric vehicles is small yet growing. We can be industry leaders in how we produce and use electric vehicles. A variety of different strategies should be employed to encourage their uptake. We should widen accessibility in the use of electric vehicles to make them more practical for those living in urban or built-up areas. For instance, by making sure there is sufficient charging infrastructure. If we use electric vehicles, journeys can be greener and safer.

Treatment number	Treatment text
Treatment 16: Aggression, Treatment, Transport	**[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.]** 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. **[People are choking to death because we are failing to clean up the toxic air we breathe. It is shameful and nothing has been done. Why?]** 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. **[Things must change, and anyone who opposes this cannot claim they care about the well-being of our people.]**
Treatment 17: Evidence, Statistics, Transport	Transport is the largest carbon-emitting sector of the UK economy, and, within this, cars contribute most. **[Last year, the transport sector accounted for 29.8% of total carbon dioxide emissions.]** Air pollution increases the risk of heart disease, cancer, diabetes, and asthma attacks. **[In London alone, there are 9,400 premature deaths every year because of poor air quality.]** Electric vehicles offer one method of reducing emissions as they produce no air pollution. The market for electric vehicles is small yet growing. **[Last year saw the biggest ever annual increase in electric vehicles, with a growth of 175,000 new vehicles, which was up 66% on 2019.]** We can be industry leaders in how we produce and use electric vehicles. 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.
Treatment 18: Evidence, Anecdote, Transport	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.

Full models

In this section, I present the full regression results from the analysis presented in the main body of the paper.

Unconditional effects

First, I the results from the analysis investigating the unconditional effects between style usage and style perceptions, likeability evaluations, and competence evaluations in tables S5–S7.

Table S5: Relationship between style usage and style perceptions

	Dependent variable:		
	Perceived emotion	Perceived aggression	Perceived evidence
	(1)	(2)	(3)
Intercept	3.242***	2.918***	3.686***
	(0.035)	(0.041)	(0.037)
Emotional	0.598***		
	(0.050)		
Aggressive		0.416***	
		(0.058)	
Anecdotal			−0.253***
			(0.052)
Observations	1,590	1,587	1,557
R^2	0.083	0.031	0.015
Adjusted R^2	0.082	0.031	0.014
Note:			*p<0.1; **p<0.05; ***p<0.01

Table S6: Relationship between style usage and likeability evaluations

	Dependent variable:		
	Likeability		
	Emotion	Aggression	Evidence
	(1)	(2)	(3)
Intercept	3.430***	3.363***	3.380***
	(0.030)	(0.032)	(0.028)
Emotional	0.029		
	(0.042)		
Aggressive		−0.271***	
		(0.045)	
Anecdotal			0.125***
			(0.039)
Observations	1,487	1,492	1,501
R^2	0.0003	0.024	0.007
Adjusted R^2	−0.0004	0.023	0.006
Note:		*p<0.1; **p<0.05; ***p<0.01	

Table S7: Relationship between style usage and competence evaluations

	Dependent variable:		
	Competence		
	Emotion	Aggression	Evidence
	(1)	(2)	(3)
Intercept	3.624***	3.615***	3.641***
	(0.031)	(0.032)	(0.031)
Emotional	−0.208***		
	(0.044)		
Aggressive		−0.114**	
		(0.046)	
Anecdotal			−0.093**
			(0.043)
Observations	1,538	1,539	1,537
R^2	0.015	0.004	0.003
Adjusted R^2	0.014	0.003	0.002
Note:		*p<0.1; **p<0.05; ***p<0.01	

Conditional effects by MP gender

Second, I present the results from the analysis investigating the conditional effects by MP gender in tables S8–S10. The output from model 1 in table S8 below shows the full regression output for the relationship between MP gender, MP-likeability evaluations, and additional control variables as displayed below for the emotion style, and model 2 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The output from model 3 shows the full regression output for the relationship between MP gender, MP-likeability evaluations, and additional control variables as displayed below for the aggression style, and model 4 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The output from model 5 shows the full regression output for the relationship between MP gender, MP-likeability evaluations, and additional control variables as displayed below for the evidence style, and model 6 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The treatment styles are the emotional style, aggressive style, and anecdotal evidence, the control styles are the non-emotional style, non-aggressive style, and statistical evidence.

In all three models for the likeability outcomes, I have chosen to include respondent gender, age, and degree education as pre-treatment covariates, as, in expectation, they will increase the statistical power of the analysis by better explaining the variation in the likeability outcomes. For instance, a respondent's gender may explain how likeable they deem a politician to be, and, indeed, the results from the conditional analysis by respondent gender presented in tables S15–S17 suggest that this is the case. Similarly, a respondent's age may provide a plausible explanation for why they would evaluate a politician delivering, say, an aggressive argument as more or less likeable.

The output from model 1 in table S9 below shows the full regression output for the relationship between MP gender, MP-competence evaluations, and additional control

Table S8: Relationship between MP gender and MP-likeability evaluations (treatment styles are emotional, aggressive, and anecdotal. Control styles are non-emotional, non-aggressive, and statistical)

	Dependent variable:					
	Likeability					
	Emotion		Aggression		Evidence	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.651***	3.592***	3.668***	3.525***	3.420***	3.345***
	(0.085)	(0.092)	(0.094)	(0.099)	(0.081)	(0.085)
Woman MP	−0.040	−0.038	0.107*	0.105*	0.082	0.086
	(0.059)	(0.059)	(0.063)	(0.063)	(0.054)	(0.054)
Treatment Style	−0.034	−0.033	−0.221***	−0.220***	0.151***	0.148***
	(0.057)	(0.057)	(0.062)	(0.062)	(0.055)	(0.055)
Woman Voter	0.188***	0.191***	0.146***	0.145***	0.273***	0.270***
	(0.042)	(0.042)	(0.045)	(0.045)	(0.039)	(0.039)
Age	−0.006***	−0.006***	−0.008***	−0.008***	−0.005***	−0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Degree Educated	0.018	0.017	−0.085*	−0.070	0.046	0.041
	(0.045)	(0.045)	(0.048)	(0.048)	(0.042)	(0.041)
Housing Policy		0.077		0.242***		0.136***
		(0.051)		(0.056)		(0.045)
Health Policy		0.078		0.164***		0.115**
		(0.051)		(0.053)		(0.048)
Woman MP*Treatment Style	0.136*	0.133	−0.132	−0.133	−0.043	−0.039
	(0.082)	(0.082)	(0.088)	(0.088)	(0.077)	(0.076)
Observations	1,487	1,487	1,492	1,492	1,501	1,501
R^2	0.037	0.039	0.059	0.072	0.058	0.065
Adjusted R^2	0.033	0.034	0.056	0.067	0.054	0.060
Note:						*p<0.1; **p<0.05; ***p<0.01

variables as displayed below for the emotion style, and model 2 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The output from model 3 shows the full regression output for the relationship between MP gender, MP-competence evaluations, and additional control variables as displayed below for the aggression style, model 4 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The output from model 5 shows the full regression output for the relationship between MP gender, MP-competence evaluations, and additional control variables as displayed below for the evidence style, and model 6 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The treatment styles are the emotional style, aggressive style, and anecdotal evidence, the control styles are the non-emotional style, non-aggressive style, and statistical evidence.

As with the likeability outcomes, I have selected pre-treatment covariates that may plausibly explain variation in respondent's competency evaluations. These are respondent gender, age, degree education, and political attention. For instance, political attention has been included as a pre-treatment covariate as respondent's with higher levels of political attention may be better equipped to judge the competence of politicians delivering arguments on political issues than respondents with low political attention. Similarly, education has been included as a voter's education level may plausible explain how competent they deem politicians to be.

The output from model 1 in table S10 below shows the full regression output for the relationship between MP gender, perceived emotion, and additional control variables as displayed below for the emotion style, and model 2 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The output from model 3 shows the full regression output for the relationship between MP gender, perceived aggression, and additional control variables as displayed below for the

Table S9: Relationship between MP gender and MP-competence evaluations (treatment styles are emotional, aggressive, and anecdotal, control styles are non-emotional, non-aggressive, and statistical)

	Dependent variable:					
	Emotion		Competence Aggression		Evidence	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.800***	3.630***	3.669***	3.552***	3.593***	3.477***
	(0.107)	(0.111)	(0.114)	(0.119)	(0.108)	(0.110)
Woman MP	−0.040	−0.035	0.020	0.016	0.075	0.083
	(0.060)	(0.060)	(0.063)	(0.063)	(0.060)	(0.060)
Treatment Style	−0.267***	−0.263***	−0.111*	−0.112*	−0.052	−0.059
	(0.060)	(0.059)	(0.063)	(0.063)	(0.061)	(0.061)
Woman Voter	0.295***	0.301***	0.217***	0.218***	0.278***	0.271***
	(0.044)	(0.044)	(0.047)	(0.047)	(0.044)	(0.044)
Age	−0.008***	−0.007***	−0.007***	−0.007***	−0.006***	−0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Degree Educated	−0.036	−0.038	−0.024	−0.010	0.054	0.044
	(0.048)	(0.047)	(0.050)	(0.050)	(0.048)	(0.047)
Political Attention	0.011	0.010	0.026**	0.026**	0.020**	0.020**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Housing Policy		0.191***		0.192***		0.240***
		(0.053)		(0.057)		(0.050)
Health Policy		0.274***		0.142***		0.152***
		(0.053)		(0.054)		(0.053)
Woman MP*Treatment Style	0.124	0.113	−0.039	−0.035	−0.077	−0.068
	(0.085)	(0.084)	(0.090)	(0.090)	(0.085)	(0.084)
Observations	1,538	1,538	1,539	1,539	1,537	1,537
R^2	0.070	0.087	0.038	0.046	0.048	0.063
Adjusted R^2	0.066	0.081	0.034	0.040	0.044	0.058

Note: *p<0.1; **p<0.05; ***p<0.01

aggression style, and model 4 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. The output from model 5 shows the full regression output for the relationship between MP gender, perceived evidence, and additional control variables as displayed below for the evidence style, and model 6 shows the non-pre-registered regression output with the inclusion of additional controls for policy area. In all three models, the treatment styles are the emotional style, aggressive style, and anecdotal evidence. The control styles are the non-emotional style, the non-aggressive style, and statistical evidence.

Finally, as with the above models, I have again included pre-treatment covariates that may plausibly explain respondents' style perceptions. For emotion and aggression, these are respondent gender, left-right placement, and age. For evidence, this is respondent gender, age, and degree education. Left-right placement has been included in the models for perceived emotion and aggression as work has suggested that left-wing ideology has historically been associated with greater emotionality than right-wing, and therefore voters' left-right positioning may explain variation in the emotion and aggression perceptions (e.g. Salmela and Von Scheve, 2018). Similarly, I include degree education in the perceived evidence model, as a respondent's degree education may plausibly explain how they evaluate the extent to which they perceive anecdotal or statistical arguments to be evidence-based.

Table S10: Relationship between MP gender and style perceptions (treatment styles are emotional, aggressive, and anecdotal. Control styles are non-emotional, non-aggressive, and statistical)

	Dependent variable:					
	Perceived emotion		Perceived aggression		Perceived evidence	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.144***	2.905***	2.768***	2.714***	3.606***	3.468***
	(0.125)	(0.135)	(0.150)	(0.157)	(0.111)	(0.115)
Woman MP	−0.107	−0.099	0.088	0.082	0.041	0.049
	(0.081)	(0.081)	(0.095)	(0.095)	(0.074)	(0.073)
Treatment Style	0.679***	0.692***	0.567***	0.565***	−0.206***	−0.213***
	(0.081)	(0.080)	(0.094)	(0.094)	(0.075)	(0.075)
Woman Voter	0.065	0.070	−0.091	−0.089	0.220***	0.212***
	(0.058)	(0.058)	(0.068)	(0.068)	(0.053)	(0.053)
Left-Right Placement	0.013	0.022	−0.012	−0.012		
	(0.021)	(0.021)	(0.025)	(0.025)		
Age	0.00004	0.0002	0.003	0.003	−0.001	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Housing Policy		0.187***		0.059		0.278***
		(0.072)		(0.084)		(0.061)
Health Policy		0.352***		0.110		0.185***
		(0.072)		(0.080)		(0.065)
Degree Educated					−0.045	−0.056
					(0.056)	(0.056)
Woman MP*Treatment Style	0.067	0.048	−0.174	−0.168	−0.081	−0.071
	(0.116)	(0.115)	(0.135)	(0.135)	(0.104)	(0.103)
Observations	1,226	1,226	1,220	1,220	1,557	1,557
R^2	0.112	0.129	0.044	0.046	0.028	0.041
Adjusted R^2	0.107	0.123	0.040	0.040	0.024	0.036
Note:						*p<0.1; **p<0.05; ***p<0.01

Multiple comparisons correction

In the empirical strategy carried out in the main body of the paper, there are numerous statistical tests are conducted, and there is risk of the multiple comparisons problem. To quote Gelman, Hill and Yajima (2012, 189-190), the multiple comparisons problem "is that the probability of a researcher wrongly concludes that there is at least one statistically significant effect across a set of tests, even when in fact there is nothing going on, increases with each additional test". In other words, when conducting a very large number of tests any one might be "significant" by chance alone, meaning that the p-values are unlikely to capture the true Type 1 error rate. Therefore, the probability of falsely rejecting a null hypothesis which is correct increases with each additional test.

To assess whether the results I present in the main body are robust, I here carry out subsequent analyses with adjusted p-values which control for the False Discovery Rate using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995). An alternative common approach is the Bonferroni correction. Under this approach, the p-value at which a test is evaluated is based on the total number of tests performed. Practically speaking, the p-value is calculated as the original p-value divided by the number of tests performed. However, the Bonferroni correction assumes independence between tests conducted which is clearly inappropriate in the case of this experimental design, where any form of gender-bias driving perceptions of women's use of, say, aggression is also likely to inform how respondents perceive and evaluate women's use of, say, statistical evidence. Further, by targeting the Type 1 error problem, the Bonferroni correction increases the likelihood of Type 2 errors. By changing the p-value needed to reject the null hypothesis, it increases the number of instances where the null is not rejected when it is in fact false and should have been. As Gelman, Hill and Yajima (2012, 192) argue: "the Bonferroni correction can severely reduce our power to detect an important effect".

I instead opt to use the Benjamini-Hochberg (Benjamini and Hochberg, 1995) procedure which is less stringent than the Bonferroni correction and is more appropriate in the context of this experimental design. In practice, when using the Benjamini–Hochberg procedure, an α level is selected, which here is 0.05. Next, the p-values of all hypotheses tests are ordered. Then I identify the largest p-value which satisfies the criteria $p_k \leq \frac{k}{m}\alpha$ where k is the p-value's index. This test, and all tests with smaller p-values are declared significant. I apply this procedure to all of the unconditional and conditional models run in the main body of the text. The relevant p-values in the unconditional models are those that relate to the differences between treatment and control styles – β_1. The relevant p-values in the conditional models are those that relate to the differences in the effect between treatment and control styles for men and women MPs – β_3.

Table S11 shows the results from this procedure where I report the unadjusted and Benjamini–Hochberg adjusted p-values for the coefficients described above, whether they are significant or not, and whether the correction changes the significance. As is clear from the table, the results presented in the main body of the paper are robust to the correction in all instances except for the model showing the unconditional relationship between MP's use of evidence and competence evaluations (unconditional evidence competence). After the multiple comparisons correction is applied, this coefficient is no longer statistically significant.

Table S11: Comparison between unadjusted and Benjamini-Hochberg adjusted p-values

Model name	Unadjusted p-value	Significant?	Adjusted p-value	Significant?	Difference?
Unconditional emotion perception	0.00	Yes	0.00	Yes	No
Unconditional aggression perception	0.00	Yes	0.00	Yes	No
Unconditional evidence perception	0.00	Yes	0.00	Yes	No
Unconditional emotion likeability	0.49	No	0.59	No	No
Unconditional aggression likeability	0.00	Yes	0.00	Yes	No
Unconditional evidence likeability	0.00	Yes	0.00	Yes	No
Unconditional emotion competence	0.00	Yes	0.00	Yes	No
Unconditional aggression competence	0.01	Yes	0.03	Yes	No
Unconditional evidence competence	0.03	Yes	0.07	No	Yes
Conditional emotion likeability	0.10	No	0.20	No	No
Conditional aggression likeability	0.14	No	0.24	No	No
Conditional evidence likeability	0.58	No	0.61	No	No
Conditional emotion competence	0.15	No	0.24	No	No
Conditional aggression competence	0.67	No	0.67	No	No
Conditional evidence competence	0.37	No	0.51	No	No
Conditional emotion perception	0.56	No	0.61	No	No
Conditional aggression perception	0.20	No	0.29	No	No
Conditional evidence perception	0.44	No	0.56	No	No

Addditional non-pre-registered analyses

Pooled models

In this section, I present the analysis of two sets of analysis where I pool the style types together. To do so, I first create a variable for whether a style is **female stereotype-congruent**, which takes the value of 0 for female stereotype-incongruent styles (non-emotional style, aggressive style, and statistical evidence) and 1 for female stereotype-congruent styles (emotional style, non-aggressive style, and anecdotal evidence). To assess how politicians using styles which are congruent with female stereotypes affecting voters' likeability and competence evaluations of them. To do so for the likeability outcomes, I pool all styles together and estimate the following:

$$Likeability_{i(j)} = \alpha + \beta_1 FemaleStereotypeCongruent_j + \epsilon_i \quad \text{(S1)}$$

where α describes the average likeability evaluations in female stereotype incongruent styles (non-emotional style, aggressive style, and statistical evidence), and $\alpha + \beta_1$ describes the same quantities for female stereotype congruent styles (emotional style, non-aggressive style, and anecdotal evidence).

I can, of course, identify the same quantity for the competence outcomes. I therefore also estimate:

$$Competence_{i(j)} = \alpha + \beta_1 FemaleStereotypeCongruent_j + \epsilon_i \quad \text{(S2)}$$

where α describes the average competence evaluations in female stereotype incongruent styles (non-emotional style, aggressive style, and statistical evidence), and $\alpha + \beta_1$ describes the same quantities for female stereotype congruent styles (emotional style, non-aggressive style, and anecdotal evidence). As there are multiple observations per

respondent, I cluster standard errors in both models at the respondent level. The results are presented in table S12.

Table S12: Relationship between female stereotype congruent styles, likeability and competence evaluations

	Dependent variable:	
	Likeability	Competence
	(1)	(2)
Intercept	3.299***	3.589***
	(0.020)	(0.021)
Female Stereotype Congruent	0.144***	−0.062**
	(0.026)	(0.025)
Observations	4,480	4,614
R^2	0.008	0.001
Adjusted R^2	0.007	0.001
Note:	*p<0.1; **p<0.05; ***p<0.01	

In the left hand column, I show the results for the likeability outcomes. Here, the coefficient for female stereotype congruent (emotional style, non-aggressive style, and anecdotal evidence) is positive and significant, suggesting that politicians are evaluated as *more* likeable when they express styles which are congruent with "communal" stereotypes associated with women. This finding perhaps seems intuitive, given that communal stereotypes are associated with being warm, kind, emotional, and people-oriented (Eagly and Wood, 2012; Schneider and Bos, 2019).

In the right hand column, I show the results for the competence outcomes. Here, the coefficient for female stereotype congruent (emotional style, non-aggressive style, and anecdotal evidence) is negative and significant, suggesting that politicians are evaluated as *less* competent when they express styles which are congruent with "communal" stereotypes associated with women. 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 takeaway from these findings, therefore, is that style usage represents

a trade-off for politicians. While politicians gain in the likeability assessments when they use "communal" styles, they less out in their competence evaluations.

Next, in the main analysis conditional analysis by MP gender presented in the main body of the paper I analyse each style separately. I here present the results from a non-pre-registered analysis where I pool all styles together and compare each treatment back to the control arguments. To do so, I first create a categorical variable for the **style type** of an argument, which takes the values of control (0), emotion (1), aggression (2), statistics (3), and anecdote (4). With this variable in hand, I estimate the following:

$$\begin{aligned} Likeability_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 Emotion_j + \beta_3 Aggression_j + \\ \beta_4 Statistics_j + \beta_5 Anecdote_j + \beta_6 (WomanMP \cdot Emotion) + \\ \beta_7 (WomanMP \cdot Aggression) + \beta_8 (WomanMP \cdot Statistics) + \\ \beta_9 (WomanMP \cdot Anecdote) + \gamma X_{i(j)} + \epsilon_i \end{aligned} \tag{S3}$$

where β_1 describes the difference in likeability evaluations between women and men MPs in the control condition. β_2–β_5 describe the effect of each style type on MP likeability evaluations compared to the control condition α. β_6–β_9 describe the difference in the effect of each style type on likeability evaluations for women MPs compared to men. My argument is that women MPs will be rewarded for expressing styles that are congruent with female stereotypes, while they will be punished for expressing female stereotype incongruent styles. As such, I expect that the coefficients for β_6 and β_9 to be positive, and the coefficients for β_7 and β_8 to be negative. X_i represents a control for each policy area (transport, housing, and health). As there are multiple observations per respondent, I cluster standard errors at the respondent level i.

I also estimate the same model for the competence outcome:

$$\begin{aligned} Competence_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 Emotion_j + \beta_3 Aggression_j + \\ \beta_4 Statistics_j + \beta_5 Anecdote_j + \beta_6 (WomanMP \cdot Emotion) + \\ \beta_7 (WomanMP \cdot Aggression) + \beta_8 (WomanMP \cdot Statistics) + \\ \beta_9 (WomanMP \cdot Anecdote) + \gamma X_{i(j)} + \epsilon_i \end{aligned} \tag{S4}$$

where β_1–β_9 represent the same quantities as above. Again my argument is that women MPs will be rewarded for stereotype-congruent behaviour and punished for stereotypes-incongruent behaviour. As such, I expect that β_7 will be negative, and β_6, β_8, and β_9 will be positive. X_i again represents a control for each policy area (transport, housing, and health). I cluster standard errors at the respondent level i. The results from both models are presented in table S13.

Looking first at the left hand column of table S13, which shows the results for the likeability outcome, I see that the coefficient for β_1 (Woman MP) is insignificant. This suggests there is no difference in likeability evaluations between men and women MPs in the control condition. β_2 (Emotion) and β_4 (Statistics) are also both insignificance, suggesting that among men, the use of emotional and statistical styles have no statistically significant effect on voters' evaluations of men's likeability relative to the control. However, β_3 (Aggression) is negative and significant, suggesting that when men deliver aggressive arguments they are perceived as less likeable. Similarly, β_5 (Anecdote) is positive and significant, and voters therefore ascribe men higher likeability evaluations when using anecdotes compared to the control. The table also shows positive and significant coefficients for both housing and health policy areas, which suggests that voters find politicians to be more likeable in these issue areas compared to in transport, the baseline.

Table S13: Relationship between MP gender, style usage, likeability and competence evaluations

	Dependent variable:	
	Likeability	Competence
	(1)	(2)
Intercept	3.287***	3.492***
	(0.034)	(0.036)
Woman MP	0.028	−0.019
	(0.042)	(0.044)
Emotion	0.025	−0.253***
	(0.051)	(0.054)
Aggression	−0.265***	−0.106**
	(0.050)	(0.053)
Statistics	−0.039	−0.016
	(0.052)	(0.054)
Anecdote	0.110**	−0.075
	(0.051)	(0.054)
Housing Policy	0.156***	0.206***
	(0.030)	(0.031)
Health Policy	0.123***	0.190***
	(0.030)	(0.031)
Woman MP*Emotion	0.068	0.093
	(0.072)	(0.076)
Woman MP*Aggression	−0.077	−0.020
	(0.073)	(0.076)
Woman MP*Statistics	0.057	0.096
	(0.073)	(0.076)
Woman MP*Anecdote	0.009	0.026
	(0.073)	(0.076)
Observations	4,480	4,614
R^2	0.034	0.021
Adjusted R^2	0.032	0.018
Note:	*p<0.1; **p<0.05; ***p<0.01	

Interestingly, the coefficients for each of the interaction terms – β_6–β_9 in equation S3 described above – are all statistically insignificance. The results from this analysis therefore support the analysis in the main body of the paper: that although style usage influences voters' evaluations of politicians' likeability, these evaluations are *not* gendered.

Turning to the right hand column of table S13, I present the results from equation S4. β_1 (Woman MP) is again insignificant, and therefore also appears to be no difference in competence evaluations between men and women MPs in the control condition. β_2 (Emotion) and β_3 (Aggression) are, however, both negative and significant, and voters therefore evaluate men as less competent when they use emotional arguments or aggressive arguments compared to the control. Both β_4 (Statistics) and β_5 (Anecdote) are insignificant. As in the likeability model, the coefficients for housing and health are both positive and significant, suggesting that voters also ascribe higher competence evaluations to politicians delivering arguments on these issue areas compared to transport.

The coefficients for the each of the interaction terms – β_6–β_9 in equation S4 – are again statistically insignificant. There is no evidence that women MPs in particular are punished in competence evaluations when they violate stereotypes, nor are they rewarded when they conform to stereotypes, compared to men. Overall, the results from both models support the findings presented in the main paper.

MP gender and policy area

While not the main quantity of interest, it is possible that men and women MPs may receive differential evaluations in likeability and competence depending on the policy area in question. A rich body of literature has shown that women are stereotyped to be better suited to and more qualified on issues that relate to feminine communal stereotypes, while men are expected to instead have more authority on issues related to the masculine, agentic, and assertive stereotypes to which they are associated (Huddy and Terkildsen, 1993; Kahn, 1996; Lawless, 2004; McDermott, 1997; Schneider and Bos, 2019). Other work has shown that women candidates are more successful when they run on "women's issues" (Ennser-Jedenastik, 2017; Herrnson, Lay and Stokes, 2003). Once in office, women politicians introduce and advocate for legislation on "feminine" social policy issues (Schwindt-Bayer, 2006; Swers, 2002; Thomas, 1994), disproportionately participate in debates on women's issues (Bäck and Debus, 2019; Catalano, 2009), and may raise traditional women's issues in parliamentary speeches more than men (Bailer et al., 2022; Hargrave, 2022). Further, past experimental work has shown, for instance, that the policy context may have important implications for the power of gender stereotypes (Anzia and Bernhard, 2022; Holman, Merolla and Zechmeister, 2011, 2016, 2017). Further, women have been said to be more persuasive on "feminine" policy areas and men on "masculine" policy areas, although recent experimental work finds little evidence to support this (Anderson-Nilsson and Clayton, 2021; Searles et al., 2020).

In the experimental design described in the main body of the paper, I vary the policy area of the arguments, focusing on housing, health, and transport. While there is some debate in the literature on whether housing and transport policies are stereotypically "feminine" and "masculine" (Krook and O'Brien, 2012), health is an area that has commonly been associated with "feminine" stereotypes of being communal and caring (Catalano, 2009; Kittilson, 2011; Norris, 1996). As such, it may be the case that women are

evaluated as more likeable and competent than men on health as opposed to transport or housing policy issues. To assess whether this is the case, I estimate the following:

$$Likeability_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 HousingPolicy_j + \beta_3 HealthPolicy_j + \beta_4(WomanMP \cdot HousingPolicy) + \beta_5(WomanMP \cdot HealthPolicy) + \gamma X_{i(j)} + \epsilon_i \quad \text{(S5)}$$

where β_1 describes the difference in likeability between women and men MPs in the transport condition. β_2–β_3 describe the effect of housing and health policies on MP likeability evaluations compared to transport among men. β_4–β_4 describe the difference in the effect of housing and health policies compared to transport on likeability for women MPs compared to men. X_i represents a control for each style (emotion, aggression, and evidence). As there are multiple observations per respondent, I cluster standard errors at the respondent level i.

Similarly, I can estimate the same quantity for the competence outcomes:

$$Competence_{i(j)} = \alpha + \beta_1 WomanMP_j + \beta_2 HousingPolicy_j + \beta_3 HealthPolicy_j + \beta_4(WomanMP \cdot HousingPolicy) + \beta_5(WomanMP \cdot HealthPolicy) + \gamma X_{i(j)} + \epsilon_i \quad \text{(S6)}$$

where β_1–β_6 describe the same quantities described above. I present the results from both models in table S14. The coefficients that enable me to see whether voters are awarding differentially likeability and competence evaluations for men and women MPs are the coefficient for β_1 (Woman MP), β_4 (Woman MP*Housing Policy), and β_5 (Woman MP*Health Policy). As is clear from the table, all three coefficients are statistically insignificant, and, consequently, I see no evidence of differentially evaluations for men and women according to policy area.

Table S14: Relationship between MP gender, policy area, likeability and competence evaluations

	Dependent variable:	
	Likeability	Competence
	(1)	(2)
Intercept	3.339***	3.387***
	(0.033)	(0.037)
Woman MP	0.017	−0.012
	(0.044)	(0.047)
Housing Policy	0.153***	0.188***
	(0.038)	(0.042)
Health Policy	0.090**	0.171***
	(0.040)	(0.042)
Aggression	−0.215***	0.045*
	(0.025)	(0.027)
Evidence	0.007	0.089***
	(0.021)	(0.026)
Woman MP*Housing Policy	0.003	0.037
	(0.063)	(0.067)
Woman MP*Health Policy	0.067	0.036
	(0.063)	(0.065)
Observations	4,480	4,614
R^2	0.023	0.013
Adjusted R^2	0.022	0.011
Note:	*p<0.1; **p<0.05; ***p<0.01	

Heterogeneous effects by voter gender

How might (gendered) perceptions and evaluations of the styles politicians use vary by voter gender? In the pre-analysis plan, I stated I would carry out an exploratory analysis into how the treatment effects may differ by respondent characteristics. In this section, I carry out two sets of analysis to investigate how the effects may vary by voter gender. First, to assess whether men and women voters are equivalently sensitive to the styles that politicians use. Second, to assess whether men and women voters are differentially sensitive to the extent to which women politicians conform to stereotype-congruent behaviours. I carry out each below.

Conditional relationships by voter gender

Here I am interested in identifying whether men and women voters' evaluations and perceptions are equivalently sensitive to the styles politicians use. That is, might certain kinds of voters' evaluations of the likeability or competence of politicians be more affected by the styles used? I assess whether this is the case by interacting voter gender and style prevalence for each of the outcomes. Figure S5 shows the results for each outcome and style, and tables S15–S17 show the full results.

As described in the main body of the paper, there are few overall differences in how politicians' style usage influences the perceptions and evaluations of men and women voters. For emotion (table S15) none of the interaction terms are significant, as is also the case for aggression (table S16). However, for evidence (table S17) there are some differences. While men do not find politicians' use of anecdotal evidence as less competent or evidence-based than statistical evidence, women voters do. Further, while the use of anecdotal evidence improves men voters' likeability assessments of politicians relative to the use of statistical evidence, this is not the case for women voters. Therefore, 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 type.

Table S15: Relationship between voter gender and likeability evaluations, competence evaluations, and emotion perceptions

	Dependent variable:		
	Likeability	Competence	Emotion
	(1)	(2)	(3)
Intercept	3.282***	3.409***	3.167***
	(0.046)	(0.047)	(0.055)
Woman Voter	0.255***	0.373***	0.130*
	(0.060)	(0.062)	(0.072)
Emotional Style	0.074	−0.148**	0.624***
	(0.064)	(0.066)	(0.077)
Woman Voter*Emotional Style	−0.079	−0.107	−0.046
	(0.084)	(0.087)	(0.101)
Observations	1,487	1,538	1,590
R^2	0.018	0.049	0.086
Adjusted R^2	0.016	0.047	0.084
Note:			*p<0.1; **p<0.05; ***p<0.01

Table S16: Relationship between voter gender and likeability evaluations, competence evaluations, and aggression perceptions

	Dependent variable:		
	Likeability	Competence	Aggression
	(1)	(2)	(3)
Intercept	3.256***	3.497***	2.939***
	(0.049)	(0.049)	(0.063)
Woman Voter	0.188***	0.207***	−0.037
	(0.064)	(0.065)	(0.083)
Aggressive Style	−0.276***	−0.144**	0.428***
	(0.069)	(0.070)	(0.090)
Woman Voter*Aggressive Style	0.002	0.041	−0.018
	(0.091)	(0.092)	(0.118)
Observations	1,492	1,539	1,587
R^2	0.035	0.020	0.032
Adjusted R^2	0.033	0.018	0.030
Note:			*p<0.1; **p<0.05; ***p<0.01

Table S17: Relationship between voter gender and likeability evaluations, competence evaluations, and evidence perceptions

	Dependent variable:		
	Likeability	Competence	Evidence
	(1)	(2)	(3)
Intercept	3.167***	3.416***	3.489***
	(0.042)	(0.047)	(0.057)
Woman Voter	0.367***	0.382***	0.335***
	(0.055)	(0.061)	(0.075)
Anecdote	0.215***	0.028	−0.126
	(0.059)	(0.066)	(0.080)
Woman Voter*Anecdote	−0.151*	−0.200**	−0.212**
	(0.078)	(0.087)	(0.105)
Observations	1,501	1,537	1,557
R^2	0.045	0.033	0.029
Adjusted R^2	0.043	0.031	0.027
Note:			*p<0.1; **p<0.05; ***p<0.01

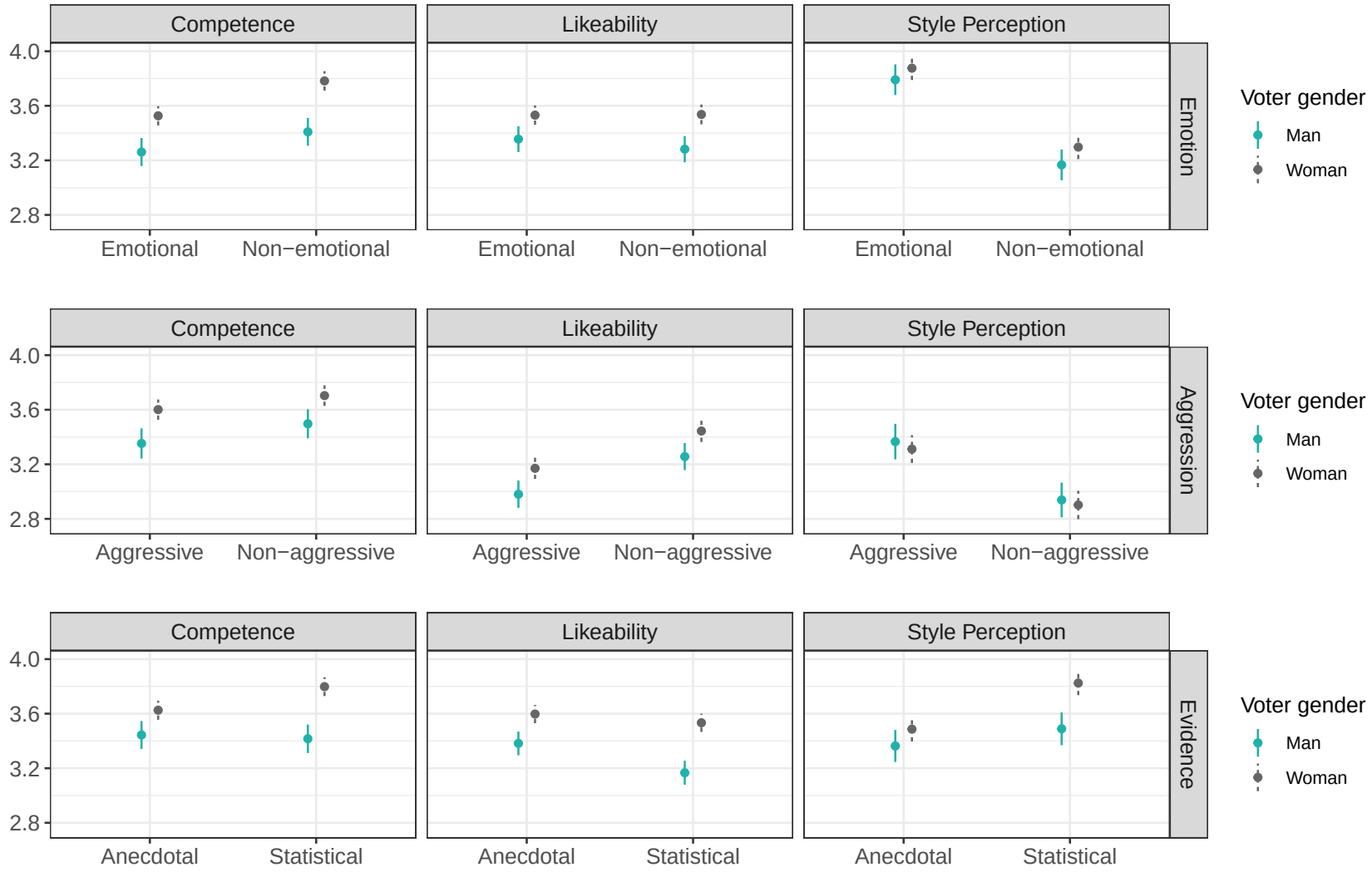

Figure S5: **Conditional relationship between voter gender, style treatment status, style perceptions, and MP likeability and competence evaluations.** The emotional style, non-aggressive style, and anecdotal evidence are female stereotype-congruent, and the non-emotional style, aggressive style, and statistical evidence are female stereotype-incongruent.

Conditional relationships by voter and MP gender

As described in the main body of the paper, I am also interested in identifying whether men and women voters are differentially sensitive to the extent to which politicians conform to or violate stereotype-congruent behaviours. Here, I assess whether this is the case by subsetting the data into men and women voters, and replicating the analysis from the main body of the paper. Tables S18–S20 present the results from each of the outcomes arranged by the style types.

While for aggression (table S19) and evidence (table S20) there do not seem to be any differences between men and women voters, this is not the case for emotion (table S18). For likeability, among women voters, I see that the coefficient on the interaction term is positive and significant, suggesting that women politicians *in particular* are rewarded for expressing emotional styles instead of non-emotional styles. I do not find the same effect among men voters. For competence, the interaction term is again positive and significant, suggesting that while women voters find emotional politicians to be less competent than non-emotional politicians, they give women MPs *more* of a competency reward than men MPs. I again see no such effect among men voters. Finally, for perceived styles I again see that the interaction term is positive and significant. While women voters perceive both men and women politicians as more emotional when they use emotional styles than non-emotional styles, they perceive women MPs *in particular* to be more emotional than men MPs. I, again, do not find the same effect among men voters.

Therefore, women voters give a bigger likeability and competence reward to women politicians who are emotional and perceive women MPs as more emotional than men MPs. I find no evidence of these effects among men respondents. Therefore, to the extent that there is any evidence of differential evaluations between men and women voters of men and women politicians for the styles they use, it seems that this effect is

concentrated amongst women voters for the emotion style type.

Table S18: Relationship between MP gender, voter gender, and likability, competence, and perceived emotion for the emotion style

	Dependent variable:					
	Likeability		Competence		Perceived emotion	
	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.284***	3.571***	3.394***	3.832***	3.103***	3.378***
	(0.067)	(0.050)	(0.072)	(0.050)	(0.079)	(0.061)
Woman MP	−0.004	−0.073	0.032	−0.105	0.134	−0.170*
	(0.098)	(0.074)	(0.104)	(0.073)	(0.115)	(0.089)
Emotional Style	0.062	−0.118*	−0.088	−0.406***	0.775***	0.442***
	(0.096)	(0.071)	(0.104)	(0.071)	(0.114)	(0.087)
Woman MP*Emotional Style	0.023	0.231**	−0.119	0.310***	−0.304*	0.285**
	(0.137)	(0.103)	(0.148)	(0.102)	(0.162)	(0.125)
Observations	620	867	647	891	665	925
R^2	0.002	0.007	0.007	0.038	0.087	0.090
Adjusted R^2	−0.003	0.003	0.003	0.035	0.083	0.087
Note:						*p<0.1; **p<0.05; ***p<0.01

Table S19: Relationship between MP gender, voter gender, and likability, competence, and perceived aggression for the aggression style

	Dependent variable:					
	Likeability		Competence		Perceived aggression	
	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.236***	3.365***	3.503***	3.685***	2.854***	2.899***
	(0.072)	(0.058)	(0.077)	(0.054)	(0.092)	(0.075)
Woman MP	0.041	0.156*	−0.012	0.038	0.170	0.006
	(0.101)	(0.082)	(0.109)	(0.077)	(0.129)	(0.106)
Aggressive Style	−0.177*	−0.215***	−0.074	−0.109	0.544***	0.494***
	(0.102)	(0.079)	(0.111)	(0.075)	(0.131)	(0.103)
Woman MP*Aggressive Style	−0.195	−0.113	−0.139	0.017	−0.232	−0.181
	(0.143)	(0.114)	(0.157)	(0.108)	(0.185)	(0.149)
Observations	627	865	641	898	667	920
R^2	0.027	0.030	0.008	0.005	0.034	0.035
Adjusted R^2	0.022	0.027	0.003	0.002	0.029	0.032
Note:						*p<0.1; **p<0.05; ***p<0.01

I present the results from this analysis graphically in figure S6. The rows represent the styles and the columns represent the outcomes. Each panel shows the four combi-

Table S20: Relationship between MP gender, voter gender, and likability, competence, and perceived evidence for the evidence style

	Dependent variable:					
	Likeability		Competence		Perceived evidence	
	Men	Women	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	3.116***	3.491***	3.362***	3.764***	3.403***	3.843***
	(0.066)	(0.048)	(0.078)	(0.050)	(0.090)	(0.065)
Woman MP	0.096	0.085	0.102	0.067	0.161	−0.036
	(0.090)	(0.068)	(0.107)	(0.070)	(0.123)	(0.091)
Anecdote	0.222**	0.098	0.084	−0.144**	−0.076	−0.293***
	(0.093)	(0.067)	(0.110)	(0.071)	(0.127)	(0.091)
Woman MP*Anecdote	−0.016	−0.066	−0.105	−0.057	−0.096	−0.095
	(0.127)	(0.096)	(0.150)	(0.100)	(0.172)	(0.129)
Observations	634	867	648	889	657	900
R^2	0.021	0.004	0.002	0.014	0.006	0.032
Adjusted R^2	0.016	0.0004	−0.003	0.011	0.002	0.029
Note:						*p<0.1; **p<0.05; ***p<0.01

nations of voter gender and MP gender, where men voters' evaluations of men politicians are displayed in the dark blue squares, men voters' evaluations of women politicians are displayed in the light blue circles, women voters' evaluations of men politicians are displayed in the triangular yellow points, and women voters' evaluations of women politicians are displayed in the maroon diamonds.

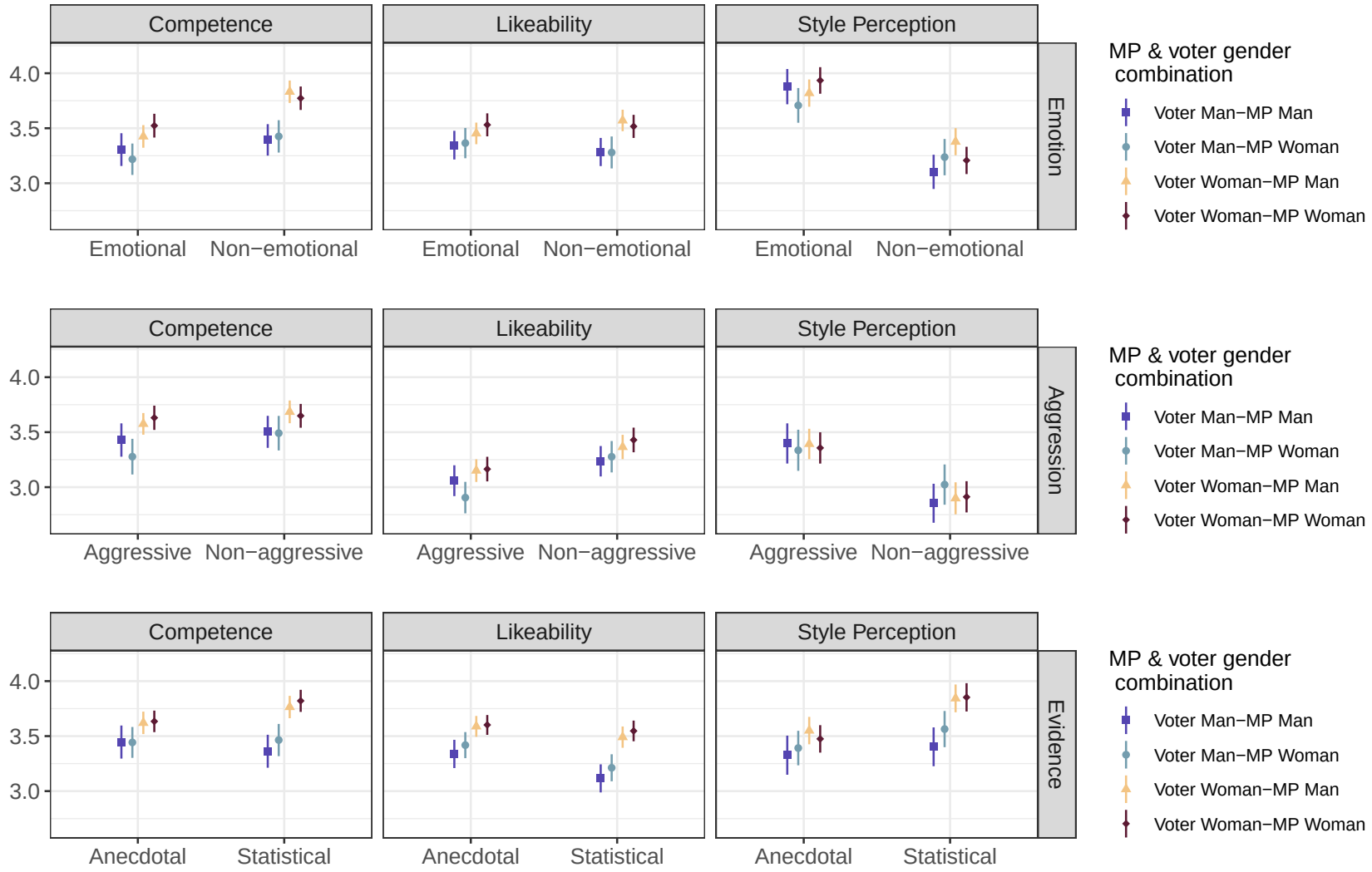

Figure S6: **Conditional relationship between MP gender, voter gender, style treatment status, style perceptions, and MP likeability and competence evaluations.** The emotional style, non-aggressive style, and anecdotal evidence are female stereotype-congruent, and the non-emotional style, aggressive style, and statistical evidence are female stereotype-incongruent.

Power analysis

Figure S7 shows the results of a power analysis for the main effects outlined in the main paper. The power analysis was conducted after the experiment and, as such, the effect size remains fixed. To construct the power analysis, I simulated the data collection process for the fixed sample of 1,600 respondents I have available for different hypothetical standardised effect sizes for the interaction between style treatment and MP gender (from very small – 0.01 standard deviations – up to large – 0.8 – standard deviations according to conventional Cohen's d standards). Note that here, to simplify the power analysis, I treat the styles as separate factorial designs, where I have a treatment and control condition and a binary moderator (MP gender) and I treat the outcome as a continuous variable, as opposed to a 5-point Likert scale.

On the basis of an analysis with 1,600 respondents – the fixed number I have available – the power analysis suggests that I am well powered to detect a standardised interaction effect size of 0.28 with 80% power. While this is reasonably small by conventional Cohen's d standards, it is also roughly comparable to the effects I estimate for the style treatments for men MPs across the various styles and outcomes. This suggests that I cannot confidently rule out the possibility of non-negligible interaction effects, but any interactions that do exist are nevertheless unlikely to be large relative to the overall variance in the outcome variables I study.

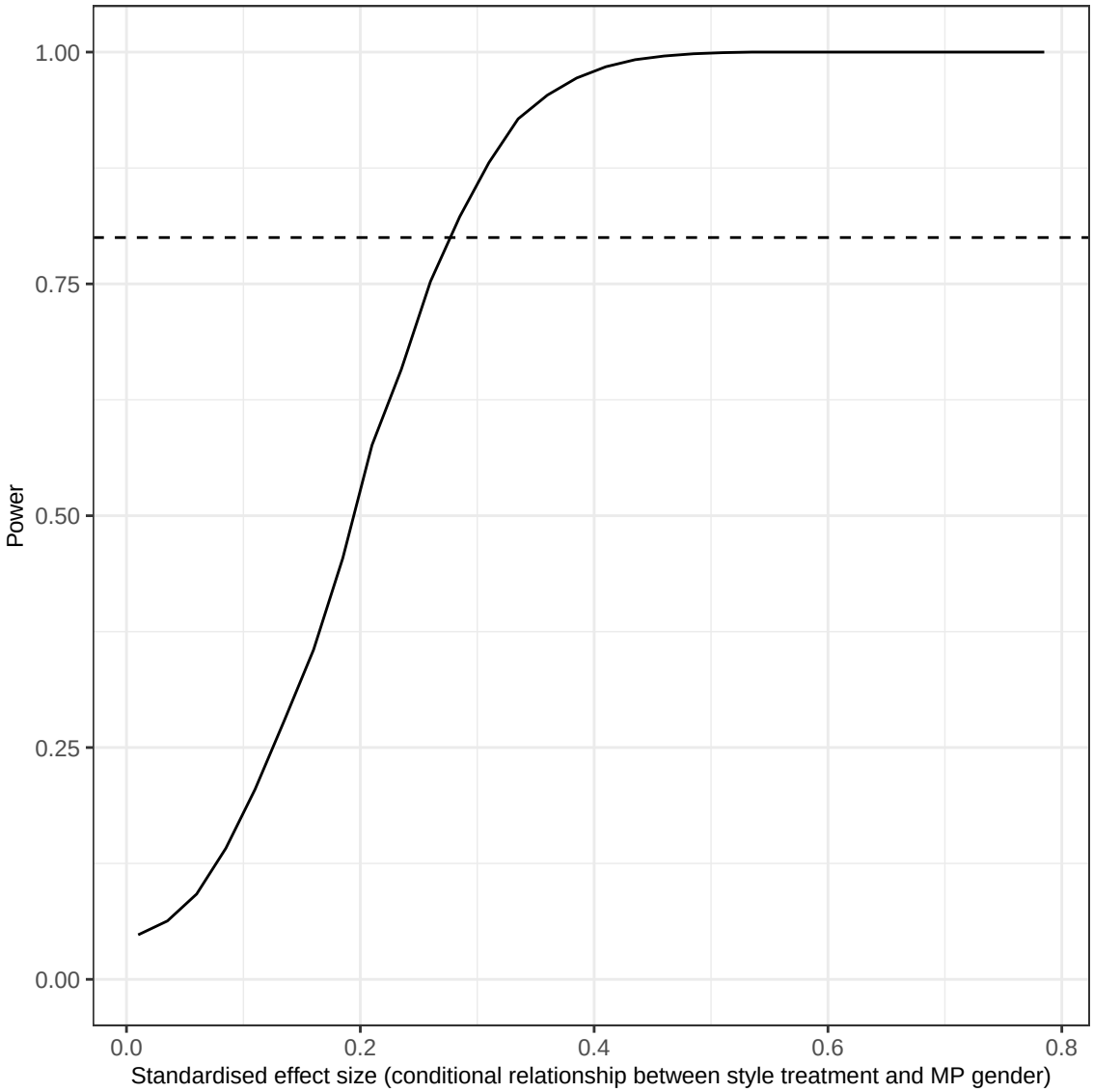

Figure S7: Power analysis

References

Anderson-Nilsson, Georgia and Amanda Clayton. 2021. "Gender and Policy Persuasion." *Political Science Research and Methods* 9(4):818–831.

Anzia, Sarah F. and Rachel Bernhard. 2022. "Gender Stereotyping and the Electoral Success of Women Candidates: New Evidence from Local Elections in the United States." *British Journal of Political Science* (Forthcoming):1–20.

Bäck, Hanna and Marc Debus. 2019. "When Do Women Speak? A Comparative Analysis of the Role of Gender in Legislative Debates." *Political Studies* 67(3):576–596.

Bailer, Stefanie, Christian Breunig, Nathalie Giger and Andreas Wüst. 2022. "The Diminishing Value of Minority Representation: Between Group Representation and Individual Career Paths." *British Journal of Political Science* 52(2):535–552.

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.

Catalano, Ana. 2009. "Women Acting for Women? An Analysis of Gender and Debate Participation in the British House of Commons 2005-2007." *Politics & Gender* 5(1):45–68.

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

Ennser-Jedenastik, Laurenz. 2017. "Campaigning on the Welfare State: The Impact of Gender and Gender Diversity." *Journal of European Social Policy* 27(3):215–228.

Gelman, Andrew, Jennifer Hill and Masanao Yajima. 2012. "Why We (Usually) Don't Have to Worry About Multiple Comparisons." *Journal of Research on Educational Effectiveness* 5(2):189–211.

Hargrave, Lotte. 2022. "Earning Their Stripes? Gender, Political Experience, and Policy Prioritisation in Parliamentary Debate." *Working Paper* pp. 1–42.

Herrnson, Paul S., J. Celeste Lay and Atiya Kai Stokes. 2003. "Women Running "as Women": Candidate Gender, Campaign Issues, and Voter-Targeting Strategies." *The Journal of Politics* 65(1):244–255.

Holman, Mirya R., Jennifer L. Merolla and Elizabeth J. Zechmeister. 2011. "Sex, Stereotypes, and Security: A Study of the Effects of Terrorist Threat on Assessments of Female Leadership." *Journal of Women, Politics & Policy* 32(3):173–192.

Holman, Mirya R., Jennifer L. Merolla and Elizabeth J. Zechmeister. 2016. "Terrorist Threat, Male Stereotypes, and Candidate Evaluations." *Political Research Quarterly* 69(1):134–147.

Holman, Mirya R., Jennifer L. Merolla and Elizabeth J. Zechmeister. 2017. "Can Experience Overcome Stereotypes in Times of Terror Threat?" *Research and Politics* 4(1):1–7.

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.

Kahn, Kim Fridkin. 1996. *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. New York, US: Columbia University Press.

Kittilson, Miki Caul. 2011. "Women, Parties and Platforms in Post-Industrial Democracies." *Party Politics* 17(1):66–92.

Krook, Mona Lena and Diana Z. O'Brien. 2012. "All the President's Men? The Appointment of Female Cabinet Ministers Worldwide." *The Journal of Politics* 74(3):840–855.

Lawless, Jennifer L. 2004. "Women, War, and Einning Elections: Gender Stereotyping in the Post-September 11th Era." *Political Research Quarterly* 57(3):479–490.

McDermott, Monika L. 1997. "Voting Cues in Low-Information Elections: Candidate Gender as a Social Information Variable in Contemporary United States Elections." *American Journal of Political Science* 41(1):270–283.

Norris, Pippa. 1996. "Women Politicians: Transforming Westminster?" *Parliamentary Affairs* 49(1):89–102.

Salmela, Mikko and Christian Von Scheve. 2018. "Emotional Dynamics of Right-and Left-wing Political Populism." *Humanity & Society* 42(4):434–454.

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.

Schwindt-Bayer, Leslie A. 2006. "Still Supermadres? Gender and the Policy Priorities of Latin American Legislators." *American Journal of Political Science* 50(3):570–585.

Searles, Kathleen, Erika Franklin Fowler, Travis N. Ridout, Patricia Strach and Katherine Zuber. 2020. "The Effects of Men's and Women's Voices in Political Advertising." *Journal of Political Marketing* 19(3):301–329.

Swers, Michele L. 2002. *The Difference Women Make: The Policy Impact of Women in Congress*. Chicago: University of Chicago Press.

Thomas, Sue. 1994. *How Women Legislate*. Oxford, UK: Oxford University Press.