

PERSISTENCE OF VOICE PITCH BIAS AGAINST POLICY DIFFERENCES *

Asli Ceren Cinar[†] Özgür Kıbrıs[‡]

August 8, 2022

Abstract

We use an online experiment to study the relative effect on voter behavior of candidate voice pitch and policy differences between candidates. We first demonstrate a strong *voice pitch bias*: in an election between candidates who are identical in every aspect but voice pitch, voters are significantly more likely to choose the one with the lower voice pitch. Voice pitch bias is higher in elections between men than women candidates. We then introduce a novel phenomenon. *Persistence of voice pitch bias* is the amount of policy difference needed to compensate for voice pitch bias. While persistence is also gender-dependent, the effect is now reversed: voice pitch bias is more persistent in elections between women than men candidates. Finally, as a possible mechanism we show that voters perceive candidates with a lower voice pitch as more competent and trustworthy. Our findings are robust against voters' gender and other socio-economic characteristics.

Keywords: voice pitch bias, persistence, voting, perception, elections, online experiment

*We thank Fikret Adaman, Nejat Anbarcı, Abdurrahman Aydemir, Florian Foos, Sara Hobolt, Özge Kemahlıoğlu, Casey Klofstad, Korhan Kocak, Mert Moral, and Emre Selçuk, as well as EPOP 2021, MPSA 2022, EPSA 2022, Oxford RAI Political Behaviour Workshop and LSE PSPE participants for comments and suggestions.

[†]Department of Government, London School of Economics; e-mail: a.cinar@lse.ac.uk

[‡]Faculty of Arts and Social Sciences, Sabancı University; e-mail: ozgur.kibris@sabanciuniv.edu.

1 Introduction

A growing body of literature has reliably established that political candidates benefit from having a lower voice pitch (as summarized in the next section). They are perceived to have higher leadership abilities. Hence, keeping everything else the same, a lower voice pitch brings in more votes. This *voice pitch bias* might be the reason why Margaret Thatcher, one of the most important political figures of the 20th century, worked with a voice coach early in her political career to lower her voice pitch.¹

But how important is the voice pitch in comparison to a candidate’s political position, the standard dimension on which voters are thought to evaluate candidates? How does the trade-off between voice pitch and policy-stance depend on the candidate’s gender, the socio-economic characteristics of the voter, and the policy dimensions on which the candidates contend? The objective of our study is to answer these questions using an experimental methodology. To the best of our knowledge, ours is the first study to approach these questions.

We focus on two policy dimensions, namely, per capita *public spending on health* and *public spending on education*, varied in a between-subjects design. Focusing on per capita public spending allows policy differences between candidates to be measured in monetary terms. We choose health and education since both are major valence issues (where an increase in the current level of public spending is perceived to be desirable) in Turkey, where we run our experiment, and both have been salient policy dimensions around the world, especially after the onset of Covid-19. Additionally, the Turkish public is predominantly satisfied (dissatisfied) from government policies on health (education), hence allowing us to control for attitudes towards government.

By integrating audio recordings into our experimental design, we diverge from traditional vignette experiments that analyze voter preferences. In our experiment, participants listen to two voice recordings (*i.e.*, candidates). On a given policy dimension, each recording declares a policy stance (*e.g.*, I will annually allocate X TL² per person for public health expenditures). To control for unobservable individual differences between candidates, the two voice recordings are obtained from the same person and digitally manipulated to either a higher or a lower pitch to create a difference of 1 Equivalent Rectangular Bandwidth (ERB)

¹Quoting Thatcher’s (1995, pg. 267) book, *The Path to Power*, “There was also the matter of my voice. ... an obvious problem for most women. ... my voice was naturally high-pitched, which can easily become grating. I had been told about this in earlier years and had deliberately tried to lower its tone. ... Gordon found me an expert who knew that the first thing to do was to get your breathing right, and then to speak not from the back of the throat but from the front of the mouth.”

²We use Turkish Liras (TL) as our experimental currency. At the time of the experiment 6.9 TL was equal to 1 USD.

(roughly 40 Hz) between them. We refer to these two recordings as the *lower-pitch (LP)* and the *higher-pitch (HP)* candidates.³ In our design, a participant is exposed to candidates of only one gender. Our experiments were carried out online. Online voice pitch experiments have been shown to produce results that are comparable to laboratory experiments [Feinberg et al., 2008].

In an election between LP and HP candidates who propose identical policies, a voter’s *voice pitch bias* can be measured via the likelihood that they will vote for the LP candidate. For an unbiased voter, this probability would be 50%, since the only difference between the two candidates – the voice pitch – would be irrelevant for them. If, however, the voter is more likely to vote for the LP candidate than can be attributed to chance, the difference in probability is a measure of the voter’s voice pitch bias. Using this measure, we corroborate the earlier empirical and experimental literature (summarized in the next section) in that the *average voter exhibits a significant voice pitch bias*. Furthermore, we find that the *average voter displays a higher voice pitch bias when evaluating a man candidate than a woman candidate*. More specifically, 69.12% of our participants vote for a man candidate with a lower voice pitch (hereafter, an *LP man*), 19.12 percentage points above 50%. For a woman candidate with a lower voice pitch (hereafter, an *LP woman*), however, this increment is only 9.63 percentage points.

In this paper, we forward that another dimension in understanding voice pitch bias is its “persistence” against other variables, such as policy differences between candidates. We define *persistence of voice pitch bias* as the amount of policy difference between LP and HP candidates as a result of which the voter will be indifferent between them, or equivalently, the marginal rate of substitution of voice pitch for policy in the voter’s preferences.⁴ For example, suppose candidate HP_M needs a 200 TL policy difference to offset the voice pitch bias against him whereas candidate HP_W needs a 1,000 TL policy difference. Comparing the two cases, we then conclude that the voice pitch bias is five times more persistent for candidate HP_W than it is for HP_M .

The above numbers are not arbitrary, but happen to belong to the high-pitched men (HP_M) versus women (HP_W) candidates in our study, respectively. That is, *the voice pitch*

³In line with the literature (e.g. Klostad et al. [2012] and Tigue et al. [2012]) and as confirmed with our pre-tests, the voice pitch difference between the two recordings is sufficient to create on the listener the impression that the two voices do not belong to the same person.

⁴To state more formally, a voter whose utility depends on the policy choice (p) and voice pitch (v) of a candidate will have a utility function of $U(p, v)$ and the relative importance of the voice pitch will be captured in the marginal rate of substitution between the two: $\frac{\partial U / \partial v}{\partial U / \partial p} = \frac{\Delta p}{\Delta v}$. As explained in the previous paragraph, the voice pitch difference between LP and HP candidates is 1 ERB, that is, $\Delta v = 1$. Hence, in our setting an equivalent definition of persistence is *the amount of policy difference between LP and HP candidates needed to offset voters’ voice pitch bias*.

bias voters display is, on average, five times more persistent when evaluating women as opposed to men candidates. Hence, a man candidate disadvantaged due to voice pitch bias needs a much smaller policy adjustment than a woman candidate in a similar situation.

As two potential mechanisms for voice pitch bias, we also measure how a candidate’s voice pitch affects the participants’ perception of her/his competence and trustworthiness. Competence is closely related to social dominance and the literature (discussed in the next section) shows that a candidate with lower voice pitch is perceived as more competent. On the other hand, previous findings on trustworthiness, another important characteristic for a candidate, are mixed. We contribute to this discussion by showing that a lower voice pitch creates on our participants a perception of both higher competence and higher trustworthiness. Furthermore, the effect is stronger for voters evaluating men candidates. In Section 4 we discuss how these findings might provide an underlying mechanism as to voice pitch bias.

Our study has important implications for gender politics. The female voice pitch typically ranges between 165 Hz to 255 Hz while the male voice has typically a much lower pitch, with a range from 85 Hz to 180 Hz. As a result, signals of social dominance and trust perceived from the voice pitch leverages men over women. Even when the two candidates are of the same gender, the effect of voice pitch on perception is prominent, for example favoring women candidates with more masculine traits. Elections between men candidates suffer from a similar gender bias, and as discussed above, sometimes to a higher extent. Furthermore, we theorize that policy declarations by men candidates are taken much more seriously by voters, and that, a combination of these two biases results in our two main findings, namely that (i) voters exhibit a stronger voice pitch bias when evaluating men candidates, and yet (ii) voters respond more strongly to policy differences between men candidates, making voice pitch bias more persistent for women candidates.

We also analyze how participant gender affects voting behavior. We start with the case where there are no policy differences between candidates. For men participants, we find that a higher percentage vote for the LP candidate when choosing between men than women candidates. This is not the case for women participants. Hence, we conclude that it is men participants that drive the overall finding that voice pitch bias is higher for men candidates than women candidates. On the other hand, both men and women participants strongly respond to a 200 TL policy difference between men candidates, but not between women candidates. Finally, we analyze the implications of a variety of pre-treatment covariates⁵

⁵We consider age, education level, household income, whether the participant voted in the last elections, left-right ideology, trust towards others, level of satisfaction from existing education and health policies, importance of government spending for education and health, views on whether men or women in elected office are better at handling issues with regarding to education or health, as well as participants’ completion times.

and demonstrate that our findings are robust against their inclusion in the analysis.

2 Background and Hypotheses

Voice pitch is a prominent vocal feature that has significant effect on humans’ perception of others. It can be defined as the number of vibrations per second made by the vocal folds to produce a vocalization [Tusing and Dillard, 2000]. Larger vocal folds generate lower frequencies due to slower vibrations, and hence produce lower sounding voices. Hence, the voice pitch is correlated with the signaller’s size. Additionally, the human voice pitch is sexually dimorphic. The voice pitch of an average male is almost half of that of an average female [Titze, 1994, Feinberg et al., 2005a, Vieira et al., 2015]. The literature shows that the voice pitch is perceived to signal information about physical and psychological traits such as attractiveness [Collins, 2000, Collins and Missing, 2003, Feinberg et al., 2005a,b, Jones et al., 2008], social and physical dominance [Tigue et al., 2012, Puts et al., 2007, Rezlescu et al., 2015, Schild et al., 2022], and reproductive capabilities [Feinberg et al., 2005b]. Lower voices are perceived to signal masculinity, trustworthiness, competence, and strength [Puts et al., 2007, O’Connor and Barclay, 2017, Klofstad, 2016, 2017, Banai et al., 2017, Feinberg et al., 2008, 2005b, Jones et al., 2010].

In the context of politics, a number of experiments demonstrate that candidates with lower voices receive higher votes and have a higher probability of winning elections [Anderson and Klofstad, 2012, Tigue et al., 2012, Klofstad et al., 2012, 2015, Klofstad, 2016].⁶ Banai et al. [2017] supports these findings in an empirical study of 51 presidential elections around the world. Klofstad [2016] analyzes the 2012 US House Elections and provides an overall support with an exception that will be discussed below. Some studies also analyze how voice pitch bias interacts with other variables. Laustsen et al. [2015] finds in survey experiments that voters with a more conservative stance display a higher voice pitch bias than more liberal voters. Using empirical data as well as survey experiments, Klofstad [2016] shows that older, well-educated, and politically engaged voters are the most biased in favor of candidates with lower voices. In an experimental study, Klofstad et al. [2015] establishes candidate age as an important determinant of voter choice and shows voice pitch to have an effect on perception of candidate age. In an empirical study, Klofstad and Anderson [2018] finds no correlation between a politician’s voice pitch and leadership ability.

Our paper contributes to this literature in several ways. First, we test the aforementioned finding that between two candidates who are identical in every aspect but their voice pitch,

⁶Similar results prevail outside the context of politics, for example in case of job applications [Ko et al., 2009].

a voter is more likely to vote for the one with the lower voice.

Hypothesis 1a. *Voters exhibit voice pitch bias. That is, in an election between the LP and HP candidates, voters are more likely to vote for the LP candidate.*

Second, in a novel experimental design that has not been considered before, we differentiate the two candidates on the policy space and study how voice pitch bias interacts with such differences. More importantly, we measure how much of a policy difference between the LP and HP candidates is sufficient to offset voice pitch bias. As discussed in Section 1, this amount gives us an estimate of the voter’s marginal rate of substitution between voice pitch and policy. It is what we refer to as *persistence of voice pitch bias*.

Hypothesis 1b. *By proposing a more desirable policy than her (his) opponent, a candidate can offset the vote disadvantage due to voice pitch bias.*

Gender is one of the most prominent traits voters perceive at first sight. Hence, its effects on politics have been the subject of extensive research. The literature indicates that voters believe women politicians to be warmer, more compassionate, better able to handle education, family, and women’s issues, more liberal, and feminist than men, whereas men politicians are seen as strong, intelligent, better suited to handle crime, defense, and foreign policy issues, and more conservative (see [Johns and Shephard \[2007\]](#), [Dolan \[2010\]](#) and the literature cited therein). These gender stereotypes affect voting behavior significantly. [Koch \[2000\]](#) analyzes data from the 1988–1992 Pooled Senate Election Study to show that even after candidates’ individuating ideological orientations are taken into account, candidate gender still exerts substantial effect on how voters perceive a candidate’s ideological orientation. [Eagly et al. \[2003\]](#) shows that voters are more likely to vote for candidates who endorse a position typically favored more by their own gender. [Dolan \[2010\]](#) examines how gender stereotypes shape voters’ support for women candidates in various electoral circumstances. [Johns and Shephard \[2007\]](#) finds that men voters are more inclined than women to see men candidates as stronger and to prioritise strength while voting. [Mo \[2015\]](#) analyzes how candidate quality and voter gender bias interact to determine candidate evaluation. Relatedly, [Bauer \[2020\]](#) finds that voters hold female politicians to higher qualifying requirements than male candidates. These higher standards make it more difficult for female candidates to acquire electoral support. Hence, women candidates on average are more qualified than their men counterparts. Furthermore, [Fox and Lawless \[2004\]](#) uncover that, on average, women, even those with the highest levels of professional achievement, are less likely than men to consider running for political office.

In general, women candidates face gendered constraints when running for office, and are required to “double-bind” themselves by demonstrating the competence associated with

masculinity and the tenderness associated with femininity [Carpinella and Bauer, 2021, Bauer and Santia, 2021]. Relatedly, Schneider and Bos [2014] find that women politicians do not possess the traits attributed to women (*e.g.*, warm, empathetic), and they have no advantage in terms of female-stereotypical characteristics. Using role congruity expectations as a framework, Boussalis et al. [2021] examines how candidate gender affects usage of facial, vocal, and textual communication in German federal election debates (2005–2017), as well as voters’ reaction to such communication. For example, they find that Angela Merkel expresses less anger than her male opponents, and that voters punish her for anger displays and reward her for happiness and general emotional displays. Carpinella and Bauer [2021] demonstrate that women candidates tend to blend male verbal assertions with feminine images such as presence of family, schools and hospitals.

Regarding the effect of voice pitch, Searles et al. [2017] analyzes relative effectiveness of man and woman voices in political advertising, in relation to the considered issues being masculine or feminine. Candidate and voter gender turns out to be important for voice pitch bias as well. Klofstad [2016] analyzes 2012 data on the US House Elections and finds that when facing a woman opponent, a higher voice pitch increases votes. Anderson and Klofstad [2012] finds that when considering men candidates for feminine leadership roles, women voters do not respond to voice pitch (while man voters do).

Our paper contributes to the above literature by first testing whether the amount of voice pitch bias depends on the candidate gender.

Hypothesis 2a. *Voters exhibit higher voice pitch bias when voting between men candidates than voting between women candidates.*

We also test whether candidate gender affects the marginal rate of substitution between voice pitch and policy. We hypothesize that the effect of voice pitch is more persistent in case of women candidates and hence, to offset voice pitch bias an HP woman needs to offer a much more desirable policy than an HP man.

Hypothesis 2b. *Voters exhibit a more persistent voice pitch bias when voting between women candidates than voting between men candidates.*

The literature shows voice pitch to have a strong influence on the perception of characteristics related to social power, such as competence or social dominance (*e.g.*, see Aung and Puts [2020]). Individuals with lower voices are perceived to be more competent [Klofstad et al., 2015], more socially dominant [Tigue et al., 2012, Gregory, 1994, Wolff and Puts, 2010, Borkowska and Pawlowski, 2011, Jones et al., 2010, Klofstad et al., 2015, Ko et al., 2009, Puts et al., 2007, Laustsen et al., 2015], and have better leadership abilities [Nagel

et al., 2012, Klofstad et al., 2015]. The effect of voice pitch on perceptions of trustworthiness is, on the other hand, gender-dependent. O’Connor and Barclay [2017] and Klofstad et al. [2012] find in two alternative contexts that lower pitch women voices are perceived to be more trustworthy. For man voices on the other hand, O’Connor and Barclay [2017] find that a higher pitch induces more trust, while Tigue et al. [2012] obtain an opposite finding. These contrasting findings present an interesting puzzle for us to focus on.

Our paper contributes to this discussion by testing voters’ perceptions of competence and trustworthiness of both men and women candidates when they are recorded making a policy-neutral statement “Vote for me”. In case of trustworthiness, our paper also serves to bring further evidence to the contrasting findings in the earlier literature.

Hypothesis 3a. *An LP candidate is perceived to be more competent than an HP candidate.*

Hypothesis 3b. *An LP candidate is perceived to be more trustworthy than an HP candidate.*

3 Materials and Methods

3.1 Experimental Stimuli

We recorded six native Turkish speakers - three females with an average age of 38 and three males with an average age of 40 - making the following policy statements in Turkish: ‘Please vote for me’ and ‘I will annually allocate X TL per person’. Pisanski et al. [2021] shows that studies on voice pitch obtain comparable results over different types of recordings, such as a series of vowels, a single word, or a sentence as in our case. We recorded multiple speakers to reduce any individual-level effect of other vocal characteristics such as tone of the voice, rhythm, or tempo. The monetary amount X took six values, starting from 10,000 Turkish Liras (TL) and decreasing by 200 TL at each step down to 9,000 TL. Overall, we obtained 7 recordings from each speaker.

We used monetary differences in per capita public spending to measure differences in policy. This is because monetary differences in per capita spending are easy to understand and their perception is uniform among participants (as opposed to for example differences in an abstract policy space). Also, they allow us to measure in objective units the persistence of voice pitch bias (the trade-off between voice pitch bias and policy differences).

As discussed in the previous section, we hypothesize that the LP candidate will receive significantly higher votes than the HP candidate (see Hypothesis 1a). We then hypothesize that policy differences between candidates can first mitigate and then neutralize this voice pitch bias, at which point the percentage of participants voting for the LP candidate will

not be significantly different than 50% (see Hypothesis 1b). To measure how much of a policy difference is needed to neutralize the voice pitch bias, we gradually made the policy declaration of the low-pitched candidate less desirable. After analyzing pre-test results, we determined these policy increments to be 200 TL.

To reduce any policy-specific effect on our findings, we chose two alternative policy dimensions. Our main objective in choosing education and health was to find valence issues where a higher public spending would be almost unanimously considered to be desirable. Figure 6 in the Appendix⁷ shows that this is indeed the case for our participants. Another reason for our choice of education and health was that the Turkish public is predominantly satisfied (dissatisfied) from government policies on health (education), hence allowing us to control for attitudes towards government.⁸ This is indeed corroborated by our data, as displayed in Figure 7. Hence, any finding that holds for both dimensions can not be attributed to the participants' attitude towards government. Finally, in a post-experiment survey, we asked the participants whether they think it is men or women in elected office that are better at handling each issue. As can be seen in Table 1, more than 75% of our participants express no preference and the remaining group is equally divided between preference for men and women. Hence, we believe it is appropriate for our study to consider education and health as gender-neutral issues.

To find a common monetary unit for both education and health policy statements, we consulted the OECD education and health reports for Turkey where the yearly per capita spending is stated to be approximately 2,400 TL for both dimensions. We then chose our maximal amount (10,000 TL) to be significantly higher.

Voices were recorded as .mp4 files. We inspected each audio file aurally and visually in Audacity (v.2.3.3).⁹ Before converting the audio files into .wav format, we ensured that the recordings were without speech errors and background noise. We used the *Get Pitch* command in the Praat phonetic analysis program [Boersma and Weenink, 2020, v.6.1.15] to determine the mean pitch of each recording. For unaltered woman voices, the mean pitch is 239 Hz and the standard deviation is 14 Hz. The mean pitch for unaltered man voices is 134 Hz and the standard deviation is 12 Hz. To create a lower-pitched and higher-pitched version of each recording, we used the Pitch-Synchronous Overlap Add (PSOLA) method in Praat.¹⁰ Following the literature [Jones et al., 2010, Klofstad and Anderson, 2018, Tighe

⁷All tables and figures on our descriptive statistics are in the Appendix.

⁸For example, see the 2020 **Life Satisfaction Survey** by the Turkish Statistical Institute.

⁹Audacity® software is copyright © 1999-2020 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

¹⁰This method allows us to manipulate fundamental frequency and harmonics whilst controlling the other spectrotemporal aspects of the acoustic signal [Feinberg et al., 2005a].

et al., 2012], we altered each recording by ± 0.5 Equivalent Rectangular Bandwidth (ERB). Hence, each recording was converted into a pair of recordings, one with a higher-pitch and one with a lower-pitch. The ± 0.5 ERB manipulation creates natural sounding voices and accounts for a perceivable shift of roughly ± 20 Hz. Manipulating the recordings by ERB corrects for the logarithmic difference between actual fundamental frequency and perceived fundamental frequency. Therefore, it produces a constant perceivable gap between the raised and lowered versions of a recording, regardless of its initial fundamental frequency. We ran a pre-test and 90% of our participants declared the two manipulated recordings to be different from each other.

3.2 Procedure

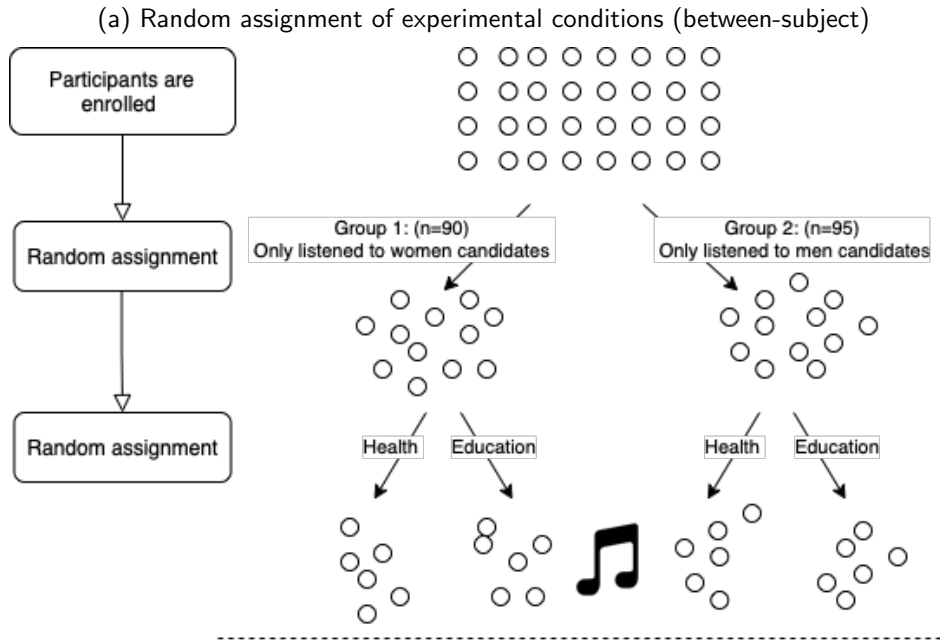
The experiment was carried out online, using the software Qualtrics. Results obtained from online voice pitch experiments have been shown to be comparable to those of laboratory experiments [Feinberg et al., 2008]. As summarized in Part (a) of Figure 1, the experimental conditions were assigned following a 2×2 factorial design (with equal probability): the participants were randomly assigned to listen to policies about either only education or only health and the candidates they were evaluating in-between were either always both men or women. Participants compared recordings of multiple speakers in order to minimize pseudoreplication bias, whereby the idiosyncratic characteristics of any one speaker might influence the results of the experiment [Kroodsma, 1990, Machlis et al., 1985].

In the first part of the experiment (Figure 1, first step of Part (b)), a participant listened to 6 individual recordings (obtained from 3 speakers) speaking the sentence ‘Please vote for me’. After listening to each recording, the participant rated the candidate in terms of trustworthiness and competence.¹¹

In the second part of the experiment (Figure 1, second step of Part (b)), a participant listened to 18 pairs of recordings obtained from 3 different speakers. After listening to each pair, the participant was asked to vote for one of the candidates. In the health policy treatment, each recording declared ‘I will annually allocate X TL per person for public health expenditures’. In each recording-pair, both recordings were obtained from the same speaker but one was higher-pitched (HP candidate) and the other one, lower-pitched (LP candidate). While the HP recording always stated $X = 10,000$ TL, the LP recording declared an X in between 9,000 TL and 10,000 TL with increments of 200 TL. The order of recording pairs, as well as, the order of recordings in each pair, were determined randomly to eliminate order effects.

¹¹Participants chose whether to use computer speakers ($n = 142$) or headphones ($n = 45$).

Figure 1: Experimental Conditions



(b) Decision tasks for each participant (within-subject, random order)

First step:

HP: "Please vote for me" vs LP: "Please vote for me"

Second step:

HP: 10,000 TL	vs	LP: 10,000 TL
HP: 10,000 TL	vs	LP: 9,800 TL
HP: 10,000 TL	vs	LP: 9,600 TL
HP: 10,000 TL	vs	LP: 9,400 TL
HP: 10,000 TL	vs	LP: 9,200 TL
HP: 10,000 TL	vs	LP: 9,000 TL

We did not inform the participants about the gender of the candidates they are listening to. Hence, any differentiation between two participants evaluating candidates of different gender is solely based on these participants' perceived gender norms regarding women and men, particularly about their voice pitch.

In the third and final part of the experiment, participants filled out a survey including questions on their birth year, sex, gender, education, family income, whether they voted in the last elections, political preferences, trust towards others, opinion on whether men or women politicians are better suited for public health/education issues, satisfaction level from public health/education services, and importance of government providing public health/education services. Participants were also asked if they had trouble listening to the recordings and the medium they used for listening (speakers versus earphones). Covariates on gender, education, age and general trust show no significant differences across the four experimental groups.

3.3 Participants

Participants ($N = 186$) predominantly declared their gender as either woman (68 participants) or man (113 participants).^{12,13} Participants ranged in age from 19 to 29 (mean age of the participants were 22 with a standard deviation of 0.12). Students who were enrolled in the introductory courses of economics received an email link that gave them access to the survey in the Qualtrics platform. Participants received course credit in exchange for their participation. They came from a diverse range of majors, belonging to the faculties of engineering, management, and social sciences. Anonymity of the participants was respected throughout the study, and their identities were kept confidential.¹⁴

Additional socioeconomic characteristics of our sample are as follows. Prior to our study, 93% of participants had voted in a real-life election. This is not significantly different than the 86% turnout rate in the 2018 Turkish elections, as published by the Turkish Statistical Institute. The monthly median family income in our sample is between 15,000 TL (2,170 USD) and 18,000 TL (2,600 USD), with a standard deviation of 10,800 TL (1,565 USD). The ideological distribution of our participants, another covariate that we are interested in, accumulates toward the mid-point, as seen in Figure 8. Furthermore, as noted in Subsection 3.1, our participants consider a higher public spending on both health and education to be desirable (Figure 6). Additionally, they are predominantly satisfied with public health

¹²4 participants preferred not to answer and 1 participant declared themselves queer.

¹³Protocols for this study were approved by Sabancı University Ethics Committee with protocol number FASS-2020-08.

¹⁴To be able to give bonus credits, course instructors were given the identification numbers of the students who participated in the study. But they did not have access to any other data.

services while the satisfaction numbers are significantly lower for education (Figure 7).

4 Results

4.1 Voice Pitch Bias

We summarize the results of our experiment in Figures 2 and 3. For each level of policy difference (taking values from 0 to 1000 on the x -axis), we conducted separate linear regressions where the randomly assigned candidate gender is the independent variable. For each participant, the dependent variable is her average vote for the LP candidate, taken over the three choice tasks (recordings).

Figure 2 provides support for *Hypothesis 1a*: when there is no policy difference between the LP and HP candidates, the percentage of participants voting for the LP candidate is 64.50%, which is significantly higher than 50% ($p < 0.01$). We also analyze subsamples where all participants vote between either all men or all women candidates. We find the percentage of participants voting for an LP man to be 69.12% ($p < 0.01$), in comparison to 59.63% for an LP woman ($p < 0.01$). Hence, in both experimental groups the probabilities are significantly higher than 50%.

Figures 2 and 3 together provide support for *Hypothesis 2a* that the magnitude of voice pitch bias responds to candidate gender. In Figure 3, when the horizontal axis is 0 TL (that is, when both candidates offer the same policy) we see a significant ($p < 0.05$) *Intention-To-Treat* (hereafter, *ITT*) effect of 9.49 percentage points. This means that the difference in votes received by an LP man and an LP woman is 9.49 percentage points. Hence, we conclude that voters exhibit a higher voice pitch bias when voting between men candidates than voting between women candidates.

4.2 Persistence of Voice Pitch Bias

Figure 2 displays – for both men and women candidates – how the percentage of participants voting for an LP candidate changes in response to the policy difference between LP and HP candidates. For both LP men and LP women, a policy difference of no more than 1,000 TL is sufficient to take this percentage down to a neighborhood of 50%, providing support for *Hypothesis 1b*. We hence conclude that by offering a more desirable policy than an LP opponent, an HP candidate can increase their vote shares and offset voice pitch bias. However, the amount of policy difference needed depends on the candidate gender, as discussed in the next paragraph.

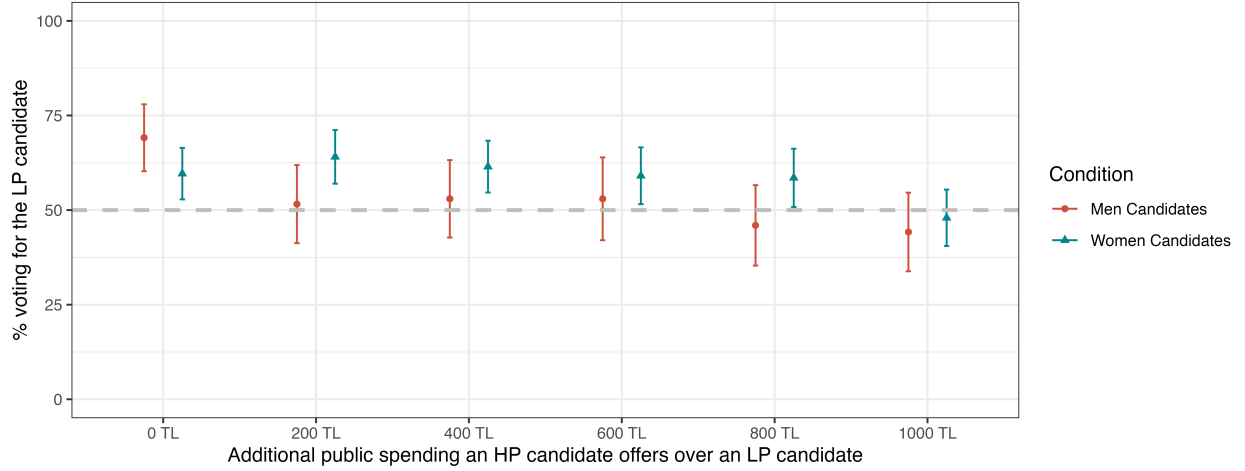


Figure 2: Percentage of participants who voted for the LP candidate by treatment condition, 95% CIs.

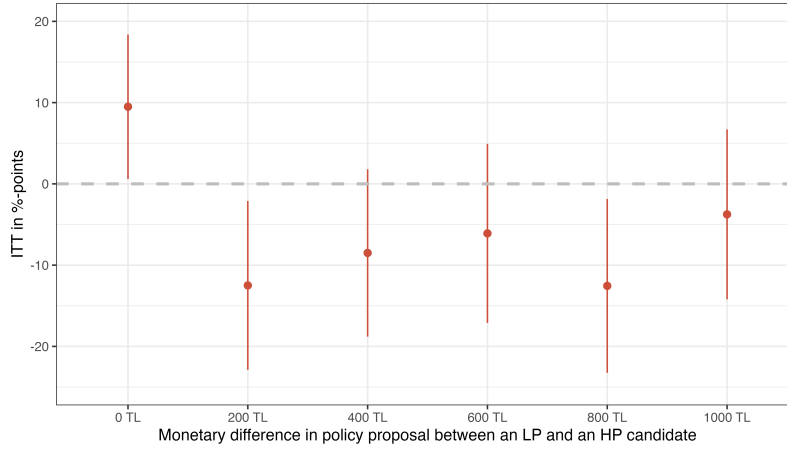


Figure 3: Effect of policy differences on the vote difference between an LP man and an LP woman.

Figure 2 shows that the percentage of participants voting for an LP man displays a sharp decrease to almost 50% as the policy difference between the candidates increases to 200 TL ($p > 0.1$).¹⁵ That is, by offering 200 TL more public spending than the LP man, the HP man is able to offset the vote disadvantages of voice pitch bias. This observation is significantly different from what we see in elections between women candidates. As can be seen in Figure 2, the percentage of participants voting for an LP woman remains significantly higher than 50% ($p < 0.05$) as the policy difference increases from 0 up to 800 TL. That is, even by offering an 800 TL more favorable policy than her opponent, an HP woman can not offset

¹⁵This p-value belongs to the one-sided t-test for the significance of the difference between the percentage votes an LP man receives at 200 TL and 50%. The other p-values in this section have a similar interpretation.

the detrimental effect of voice pitch bias on her votes. It is only when the policy difference reaches 1,000 TL that an LP and an HP woman receive more or less the same number of votes ($p > 0.1$). Overall, a comparison of men and women candidates shows us that voters exhibit a more persistent voice pitch bias when voting between women candidates than between men candidates. Hence, our findings support Hypothesis 2b.

4.3 Perceptions of Trustworthiness and Competence

Figure 4 displays the effect of a switch in voice pitch from HP to LP on perceptions of competence and trustworthiness. For candidates of each gender, we conducted separate linear regressions with the candidate voice pitch as a binary independent variable and, for each participant, her average competence (respectively trustworthiness) evaluation as a dependent variable.

Over all experimental groups, the effect on a participant’s competence rating of a switch from an HP to an LP version of the same recording is 9.43 percentage points ($p < 0.01$), supporting Hypothesis 3a. Hence, perceptions of competence provide a possible mechanism for the effect of voice pitch on voting behavior. In the two experimental groups – participants who only listened to women candidates versus participants who only listened to men candidates – the effect is 6.20 percentage points ($p < 0.05$) in case of women candidates and 12.84 percentage points ($p < 0.01$) in case of men candidates. This is in line with the higher voice pitch bias participants exhibit in case of men candidates.

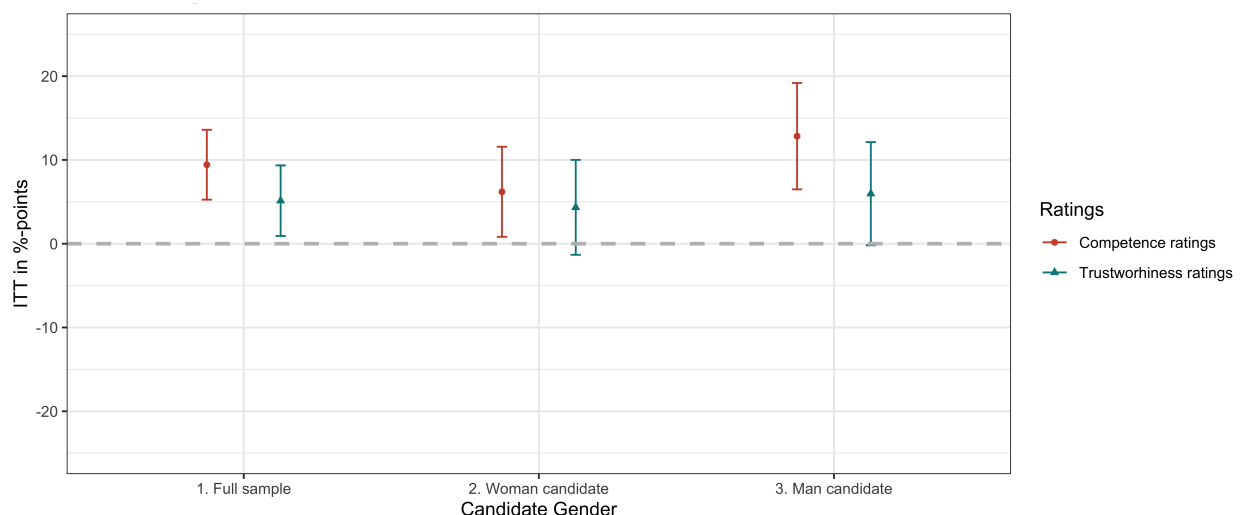


Figure 4: The effect of a switch in voice pitch from HP to LP on perceptions of competence and trustworthiness, 95% CIs.

We next analyze another theoretical mediator, trustworthiness. Figure 4 shows that, over

all experimental groups, the effect on a voter’s trustworthiness rating of a switch from an HP to an LP version of the same recording is 5.13 percentage points ($p < 0.05$), supporting Hypothesis 3b. However, for reasons that will be discussed next, trustworthiness perception seems to be a less likely mechanism underlying voice pitch bias. Namely, when the two experimental groups – participants who only listened to women candidates versus participants who only listened to men candidates – are analyzed separately, the effect on trustworthiness ratings largely diminish in significance. In case of women candidates the effect is 4.34 percentage points and not significant ($p > 0.1$). In case of men candidates, the effect is slightly larger with 5.97 percentage points, and only significant with 90% confidence ($p < 0.1$).

4.4 Exploratory Analysis: Participant Gender

In this section, we demonstrate that participant gender is an important mediator for our hypotheses on candidate gender (hypotheses 2a and 2b). Figure 5 displays the effect of participant gender on voting behavior. We first look at the case when the policy difference between candidates is 0 TL. We find that men participants vote 18.3 percentage points more for an LP candidate in elections between men than women ($p < 0.01$). In contrast, women participants vote 3.7 percentage points *less* for an LP candidate in elections between men than women, though the difference is not significant ($p > 0.1$). Hence, we conclude that it is men participants that drive the overall finding (Hypothesis 2a) that voice pitch bias is higher for men candidates than for women candidates.

For comparison, in Figure 5 we also present the case when the policy difference between candidates is 200 TL.¹⁶ Contrary to the previous case, we now find a similar treatment effect for both men and women participants. More specifically, women participants vote for an LP candidate 17.40 percentage points less in elections between men than women ($p < 0.05$). Similarly, men participants vote for an LP candidate 8.46 percentage points less in elections between men than women, though the difference is not significant ($p > 0.1$). Together with Figure 2, the above analysis suggests that for participants of both genders, a 200 TL difference between the candidates’ policies results in a significant decrease in the probability of voting for an LP man, providing a mechanism for Hypothesis 2b.

¹⁶The other cases where the policy difference is higher look very similar to the case of 200 TL. Hence, for the sake of presentation we do not show those in Figure 5.

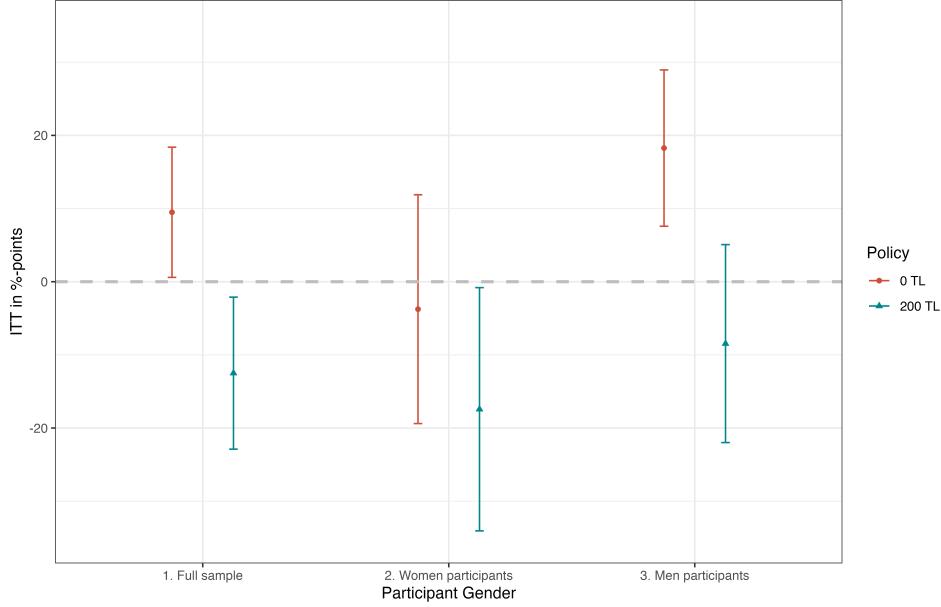


Figure 5: Effect of listening to men versus women candidates conditional on participant gender, 95% CIs.

5 Conclusion

By and large, our findings corroborate the stated hypotheses. Our participants exhibit both voice pitch bias (Hypothesis 1a) and persistence of voice pitch bias (Hypothesis 1b). Furthermore, voice pitch bias is higher in elections between men candidates (Hypothesis 2a) while it is more persistent in elections between women candidates (Hypothesis 2b). We find that men participants are mainly responsible for Hypothesis 2a while both men and women participants drive Hypothesis 2b. We also identify the effect of voice pitch on perceptions of competence (Hypothesis 3a) and trustworthiness (Hypothesis 3b) as an important mechanism for voice pitch bias.

Our results are robust to the inclusion of pre-treatment covariates, including participant gender, age, education, income level, turnout, left-right ideological stance, general trust towards others, as well as survey completion time, which is a proxy for participant attention. Additionally, our findings are consistent across the two policy dimensions – health and education – that we considered. Since the Turkish public is predominantly satisfied (dissatisfied) from government policies on health (education), we conclude that our findings are not driven by participants’ attitude towards government.

While our findings of competence and trustworthiness perceptions support Hypotheses 3a and 3b, our data also display an interesting pattern. As can be seen in Figure 4, the effect of voice pitch on perception is higher for voters evaluating men candidates. This

difference is particularly pronounced in competence ratings. Even though our participants overwhelmingly declare both education and healthcare to be gender-neutral issues (see Table 1), the earlier literature establishes them to be women-congruent [Dolan, 2014]. In relation to this literature, our study then shows that, even on woman-congruent issues, a vocal characteristic signaling masculinity – voice pitch – can have a more significant effect on perceptions regarding men than women.

References

- R. C. Anderson and C. A. Klofstad. Preference for Leaders with Masculine Voices Holds in the Case of Feminine Leadership Roles. *PLoS ONE*, 7(12):e51216, Dec. 2012.
- T. Aung and D. Puts. Voice pitch: a window into the communication of social power. *Current opinion in psychology*, 33:154–161, 2020.
- I. P. Banai, B. Banai, and K. Bovan. Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, 38(3):309–314, 2017.
- N. M. Bauer. Shifting standards: How voters evaluate the qualifications of female and male candidates. *The Journal of Politics*, 82(1):1–12, 2020.
- N. M. Bauer and M. Santia. Going feminine: Identifying how and when female candidates emphasize feminine and masculine traits on the campaign trail. *Political Research Quarterly*, page 10659129211020257, 2021.
- P. Boersma and D. Weenink. Praat: doing phonetics by computer (version 6.1.15), 2020.
- B. Borkowska and B. Pawlowski. Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1):55–59, 2011.
- C. Boussalis, T. G. Coan, M. R. Holman, and S. Müller. Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review*, 115(4):1242–1257, 2021.
- C. Carpinella and N. M. Bauer. A visual analysis of gender stereotypes in campaign advertising. *Politics, Groups, and Identities*, 9(2):369–386, 2021.
- S. A. Collins. Men’s voices and women’s choices. *Animal behaviour*, 60(6):773–780, 2000.

- S. A. Collins and C. Missing. Vocal and visual attractiveness are related in women. *Animal behaviour*, 65(5):997–1004, 2003.
- K. Dolan. The Impact of Gender Stereotyped Evaluations on Support for Women Candidates. *Political Behavior*, 32(1):69–88, Mar. 2010.
- K. Dolan. Gender Stereotypes, Candidate Evaluations, and Voting for Women Candidates: What Really Matters? *Political Research Quarterly*, 67(1):96–107, Mar. 2014.
- A. H. Eagly, A. B. Diekmann, M. C. Schneider, and P. Kulesa. Experimental Tests of an Attitudinal Theory of the Gender Gap in Voting. *Personality and Social Psychology Bulletin*, 29(10):1245–1258, Oct. 2003.
- D. Feinberg, B. Jones, A. Little, D. Burt, and D. Perrett. Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, 69(3):561–568, Mar. 2005a.
- D. R. Feinberg, B. C. Jones, L. M. DeBruine, F. R. Moore, M. J. Law Smith, R. E. Cornwell, B. P. Tiddeman, L. G. Boothroyd, and D. I. Perrett. The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, 26(5):398–408, Sept. 2005b.
- D. R. Feinberg, L. M. DeBruine, B. C. Jones, and D. I. Perrett. The Role of Femininity and Averageness of Voice Pitch in Aesthetic Judgments of Women’s Voices. *Perception*, 37(4):615–623, Apr. 2008.
- R. L. Fox and J. L. Lawless. Entering the arena? gender and the decision to run for office. *American Journal of Political Science*, 48(2):264–280, 2004.
- S. W. Gregory. Sounds of power and deference: Acoustic analysis of macro social constraints on micro interaction. *Sociological Perspectives*, 37(4):497–526, 1994.
- R. Johns and M. Shephard. Gender, candidate image and electoral preference. *The British Journal of Politics and International Relations*, 9(3):434–460, 2007.
- B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovic. Integrating cues of social interest and voice pitch in men’s preferences for women’s voices. *Biology Letters*, 4(2):192–194, 2008.
- B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovic. A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, 79(1):57–62, Jan. 2010.

- C. A. Klofstad. Candidate Voice Pitch Influences Election Outcomes: Voice Pitch Influences Voters. *Political Psychology*, 37(5):725–738, Oct. 2016.
- C. A. Klofstad. Looks and Sounds Like a Winner: Perceptions of Competence in Candidates’ Faces and Voices Influences Vote Choice. *Journal of Experimental Political Science*, 4(3): 229–240, 2017.
- C. A. Klofstad and R. C. Anderson. Voice pitch predicts electability, but does not signal leadership ability. *Evolution and Human Behavior*, 39(3):349–354, May 2018.
- C. A. Klofstad, R. C. Anderson, and S. Peters. Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, 279(1738):2698–2704, July 2012.
- C. A. Klofstad, R. C. Anderson, and S. Nowicki. Perceptions of Competence, Strength, and Age Influence Voters to Select Leaders with Lower-Pitched Voices. *PLOS ONE*, 10(8): e0133779, Aug. 2015.
- S. J. Ko, C. M. Judd, and D. A. Stapel. Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin*, 35(2):198–211, 2009.
- J. W. Koch. Do Citizens Apply Gender Stereotypes to Infer Candidates’ Ideological Orientations? *The Journal of Politics*, 62(2):414–429, May 2000.
- D. E. Kroodsma. Using appropriate experimental designs for intended hypotheses in ‘song’playbacks, with examples for testing effects of song repertoire sizes. *Animal Behaviour*, 40(6):1138–1150, 1990.
- L. Laustsen, M. B. Petersen, and C. A. Klofstad. Vote Choice, Ideology, and Social Dominance Orientation Influence Preferences for Lower Pitched Voices in Political Candidates. *Evolutionary Psychology*, 13(3):147470491560057, Sept. 2015.
- L. Machlis, P. Dodd, and J. Fentress. The pooling fallacy: problems arising when individuals contribute more than one observation to the data set. *Zeitschrift für Tierpsychologie*, 68 (3):201–214, 1985.
- C. H. Mo. The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters. *Political Behavior*, 37(2):357–395, 2015.

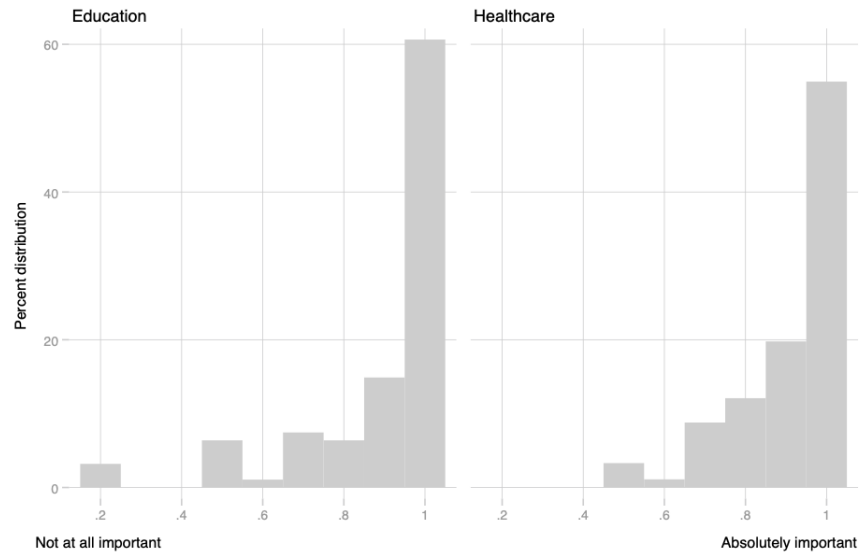
- F. Nagel, M. Maurer, and C. Reinemann. Is There a Visual Dominance in Political Communication? How Verbal, Visual, and Vocal Communication Shape Viewers' Impressions of Political Candidates: Visual Dominance in Political Communication. *Journal of Communication*, 62(5):833–850, Oct. 2012.
- J. J. O'Connor and P. Barclay. The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, 38(4):506–512, July 2017.
- K. Pisanski, A. Groyecka-Bernard, and P. Sorokowski. Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications. *Biology letters*, 17(9):20210356, 2021.
- D. A. Puts, C. R. Hodges, R. A. Cárdenas, and S. J. Gaulin. Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, 28(5):340–344, Sept. 2007.
- C. Rezlescu, T. Penton, V. Walsh, H. Tsujimura, S. K. Scott, and M. J. Banissy. Dominant Voices and Attractive Faces: The Contribution of Visual and Auditory Information to Integrated Person Impressions. *Journal of Nonverbal Behavior*, 39(4):355–370, Dec. 2015.
- C. Schild, E. Braunsdorf, K. Steffens, F. Pott, and J. Stern. Gender and context-specific effects of vocal dominance and trustworthiness on leadership decisions. *Adaptive Human Behavior and Physiology*, pages 1–19, 2022.
- M. C. Schneider and A. L. Bos. Measuring stereotypes of female politicians. *Political psychology*, 35(2):245–266, 2014.
- K. Searles, E. F. Fowler, T. N. Ridout, P. Strach, and K. Zuber. The Effects of Men's and Women's Voices in Political Advertising. *Journal of Political Marketing*, pages 1–29, May 2017.
- C. C. Tigue, D. J. Borak, J. J. O'Connor, C. Schandl, and D. R. Feinberg. Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3):210–216, May 2012.
- I. R. Titze. Control of vocal intensity and efficiency. *Principles of voice production*, 1994.
- K. J. Tusing and J. P. Dillard. The sounds of dominance. vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, 26(1):148–171, 2000.
- R. H. Vieira, C. D. Gadenz, and M. Cassol. Estudo longitudinal de caracterização vocal em canto coral. *Revista CEFAC*, 17(6):1781–1791, Dec. 2015.

6 Appendix

Table 1: Distribution of participants’ perception of gender-congruent policy domains

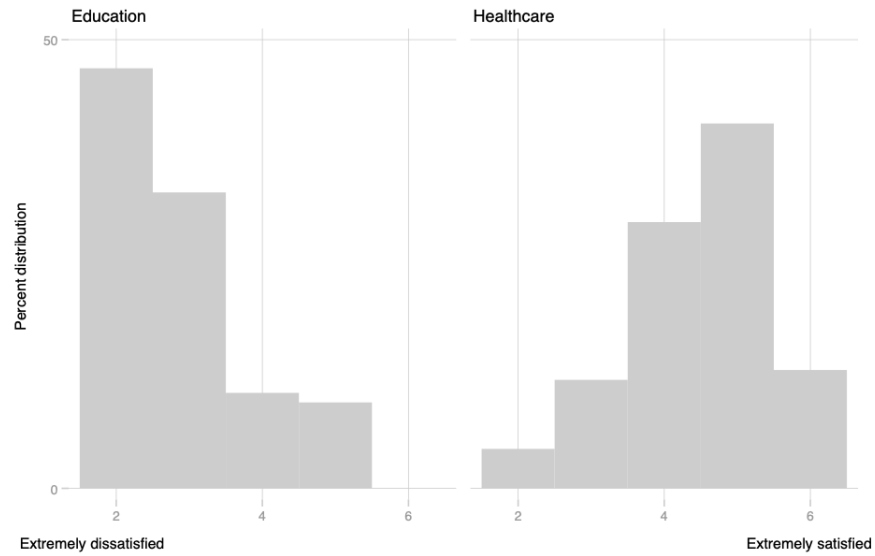
“In general, do you think men or women in elected office are better at handling issues with regard to health/education?” (N=185)				
	Men	Women	No Difference	Total
Education	8.11%	7.02%	35.68%	50.81%
Health	4.86%	3.79%	40.54%	49.19%
Total	12.97%	10.81%	76.22%	100%

Figure 6: Importance of government spending for education and health



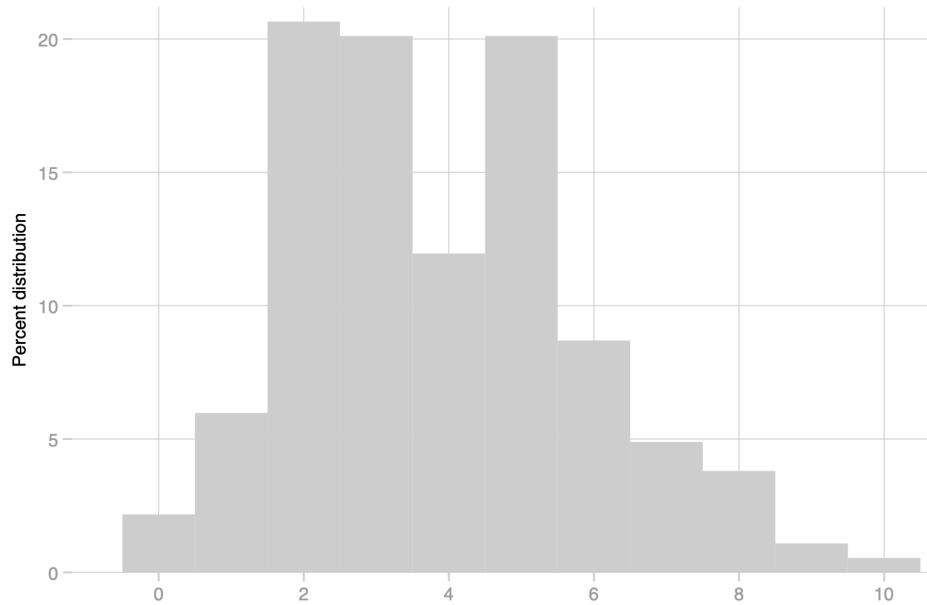
Survey Question: “How important do you think that governments provide education/health services?” (N=185)

Figure 7: Satisfaction from Education and Health Services



Survey Question: “How satisfied are you with the education/health system in Turkey?” (N=185)

Figure 8: Ideological distribution of the participants



Survey Question: “We often hear that ‘right’ and ‘left’ are mentioned in political matters. Below is a scale consisting of 10 points, with ‘1’ on the far left of the score and ‘10’ on the right. Where would you place yourself on this scale?” (N=185)