Pre-Analysis Plan: Who Bends and Who Breaks? Experimental Evidence for Conditions Affecting Voters' Evaluations of Politicians Striking Compromises

21 April, 2021 - 18:11:57 (CEST)

Contents

1 Expectations

Gender hypothesis (*H1*): Women politicians are evaluated lower when striking a political compromise compared to men politicians.

Migration hypothesis (*H2*): Politicians with a migration background are evaluated lower when striking a political compromise compared to politicians without a migration background.

Intersection hypothesis (*H3*): Women politician with a migration background are evaluated lower when striking a political compromise compared to men politicians without a migration background.

2 Research Desing and Protocol

2.1 Sample

I will conduct this survey experiment in the Netherlands in April 2021. The sample, recruited through KiesKompas, will consist of 2,000 participants (based on the power analysis presented in Figure 2) of 18 years and older. Kieskompas works with non-random opt-in respondents. Therefore, I measure many demographic background variables (see Section 3.2). Balance checks will be conducted to demonstrate whether certain categories are over represented in a certain experimental group. The study has been approved by the Research Ethics Review Committee of the *Vrije Universiteit Amsterdam* (see the approval here). To ensure good quality of our data, two attention checks (discussed in more detail in Section 3.3) are included. Each respondent failing the attention check will be excluded and replaced with another 'good' response.

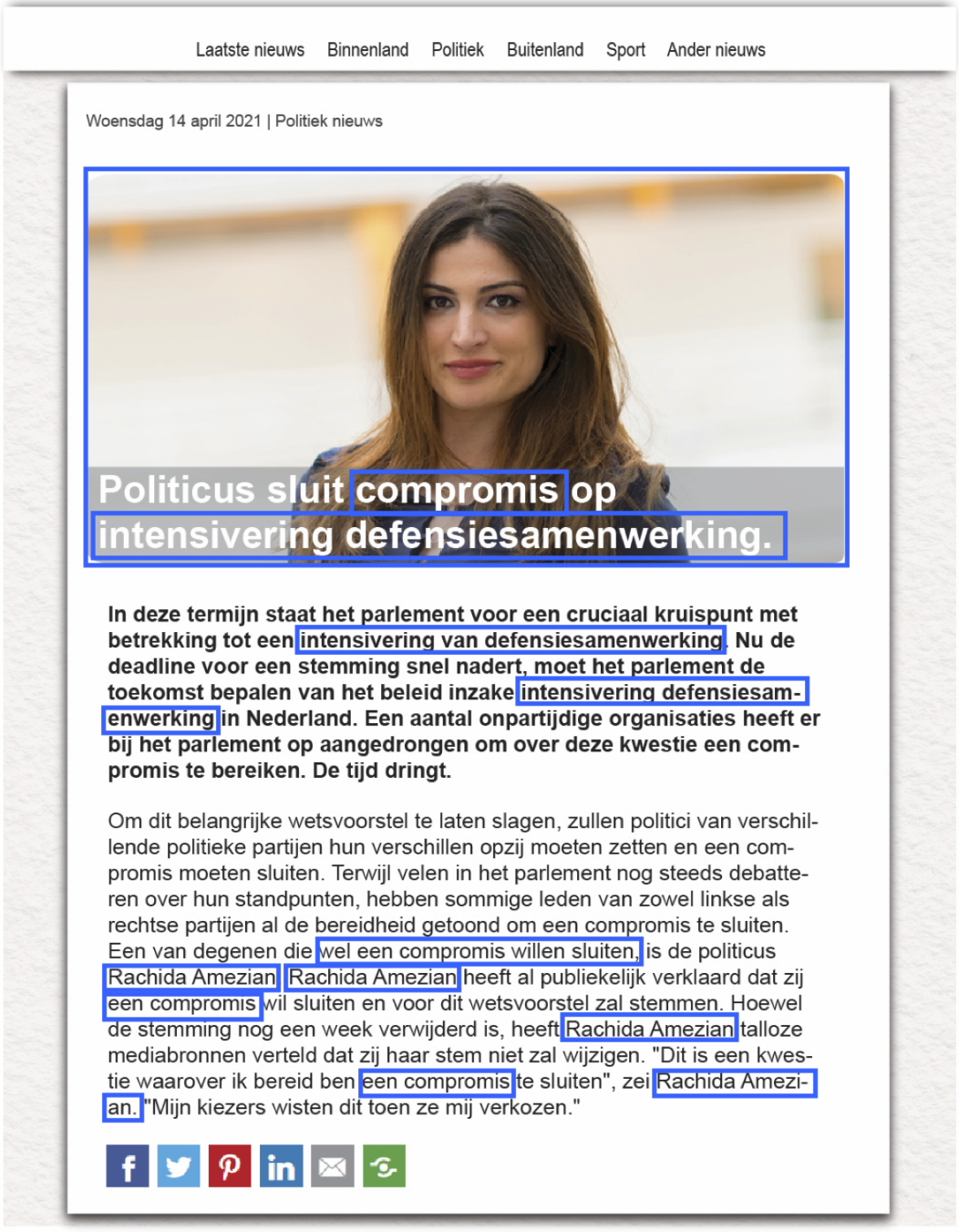
2.2 Experimental Protocol

The study is conducted online and in Dutch. Participants are told that they are taking part in a survey to get an overview of how Dutch people form their views on politics. After reading an informed consent message participants are forwarded to the main questionnaire (or the survey will be terminated if they do not agree to the consent form).

First, participants complete a set of background variables on their stances on the political issues used in the experiment (i.e. corona-measures, defense, education, immigration) and on their attitude towards women in politics – the full codeboek can be viewed here. The pre-treatment block ends with one of the two attention checks included in this survey. When participants fail this attention check, a warning appears asking them to read the question again carefully and to answer again. Only when they have answered it correctly, they enter the first round of the experiment. After each round of the experiment, some filler questions are asked about the respondents demographics. The stimuli in the experiment are news messages in the same style as the Dutch news website nu.nl. In these news messages, we manipulate: a) the *gender* of the politician (male vs. female); b) the *migration background* of the politician (based on a Arabic sounding name (Rachid(a) Amezian) or a native Dutch sounding name (Karel/Karin van der Kleijn)); and c) whether the politician *striked a compromise or not*. This creates a full `2*2*2` factorial experiment with four rounds. Every round, the news message covers a new issue. The issues are: "defense" (increase cooperation of the defense units), "immigration" (immigration stop in western countries), "education" (expansion creative elementary school activities), "COVID-19" (emergency legislation to combat COVID-19). For every round, the politicians name, and thereby gender and immigration background, are manipulated in the text and in the picture as well as the decision to strike a compromise. For an illustration of the stimulus material, see Figure 1. All manipulated elements are within the blue boxes.

The text of the treatment is a Dutch translation of the stimulus material developed by Bauer, Yong, and Krupnikov (2017). The authors have pre-tested the text of the stimuli material several times (see the Web Appendix of Bauer, Yong, and Krupnikov 2017). The photo's of their stimuli material can unfortunately not be used, because a) Bauer, Yong, and Krupnikov (2017) only investigate the difference between men and women politicians, using solely photo"s of caucasians; and b) the American flag was present in their material. Instead, I have used photo's from local politicians of the city of The Hague that are not know to the general public. The men politicians are in real life representatives of the liberal-progressive party D66 (man with

Figure 1: Annotated Example of Stimulus Material



immigrant background) and of the christian democratic party CDA (man without immigrant background). The women politicians are in real life representatives of a right-wing populist party *Hart voor Den Haag/ Groep de Mos* (woman with migration background) and of the social democratic party PvdA (woman without immigrant background). Thereby we have a spread over the ideological different parties present in Dutch politics. The names for the politicians without immigration background are *Karel* (man) and *Karin* (woman) *van der Kleijn*, which are typical native Dutch names. For the politicians with an immigration background, we picked *Rachid* (man) and *Rachida* (woman) *Amezian*.

2.3 Power Analysis

As detailed in Section 4, I conduct an OLS regression within a multiverse approach (Simonsohn, Simmons, and Nelson 2019; Steegen et al. 2016) using the four post-treatment questions on the politician's a) leadership evaluaton (scale); b) general favorability; c) representative quality; and d) career perspective as dependent variable and the three manipulations (gender, migration background and whether or not the politician is willing to strike a compromise) as independent variables. Each hypothesis is tested separately for all four issues. To calculate power for the hypotheses, the R package `DeclareDesign` is used (Blair et al. 2019). Based on the study of Bauer, Yong, and Krupnikov (2017), the effect sizes are between `b = 0.2` and `b = 0.1` – i.e. a small effect visualized by the purple and blue lines in Figure 2. The hypothesis are directional, Figure 2 therefore displays one-tailed tests with α `= 0.05`. The power analysis shows that testing hypotheses 1 and 2 requires a sample size of `1,800` participants (x-axis) to reach `95%` power (black dashed line, in the left-panel of Figure 2). Note that if the effect size is bigger then `0.2` (i.e. `0.3`, `0.4`, `0.5`, or `0.6`) smaller sample sizes are sufficient to reach `95\%` power. To test H3, the combination of gender and migration background, we can detect a small effect `b = 0.2` (blue line) with a power of `80%` (gray dashed line) and a one-tailed test of signficance at α `=0.05` with a sample of `1,800` participants (Right-Panel of Figure 2). At `2,000` participants, with a small effect and one-tailed test of significance at α `=0.05`, I reach power of `88%`. A sample of `2,000` participants therefore gives me suficient power to test Hypotheses 1 and 2, but a probability of `12%` for a Type II error remains when testing Hypotheses 3. I will therefore test all hypotheses first by issue and second, by pooling our data across issues. As Figure 3 demonstrates, this will give sufficient power even if we do two-tailed testing.

```
source(here::here("docs/pap/poweranalysis.R"))
```

3 Measures

3.1 Dependent Variables

I rely on four measures reflecting different aspects of how people judge politicians. The evaluation of politicians' character traits are statements deriving from Aaldering and Vliegenthart (2016). The other three neasures to evaluate the politician of the treatment are translated from Bauer, Yong, and Krupnikov (2017) – i.e. favorability, representation, and career perspectives. General favorability is shown to be related to vote choice (Elis, Hillygus, and Nie 2010), perceptions of represenation are related to legislative compromise and cooperation (Doherty 2013), and career perspective is included because critiques of women are typically manifested in their leadership skills (Eagly 2007). Since the questionaire is in Dutch, Table 1 describes the measures both in English and in Dutch.

1. *Evaluation of politicians' character traits* This measure exists of six statements – see Table 1 `PT-1_1` till `PT-1_6`. Each statement is measured on a 5-point Likert-scale: From `very much disagree` to `very much agree`. If `Cronbach's` α `=>0.8`, I will construct an additive scale for this measure. When `Cronbach's` α `<0.8`, I will add the statement as seperate dependent variables in the multiverse analysis.

2. *Favorability* In this measure respondents are asked to rank the politician from very unfavorable (value of `0`) to very favorable (value of `10`) for the decision to (not) strike a compromise on the issue at stake. See Table 1 `PT-2` for the exact phrasing of the question.

3. *Representation* In this measure, respondents are surveyed on how well the statement "[NAME] is a good representative of [HIS/HER] constituency" describes the politician. The politician can be evaluated on

Figure 2: Power Analysis

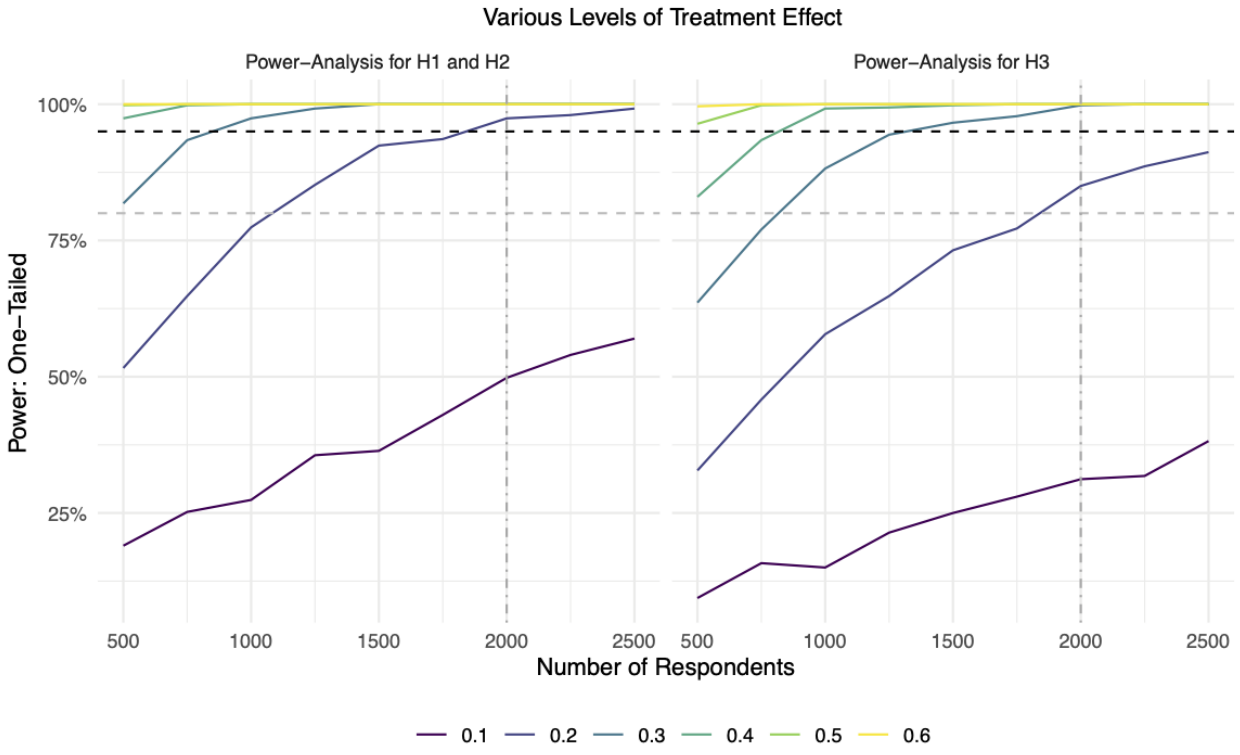
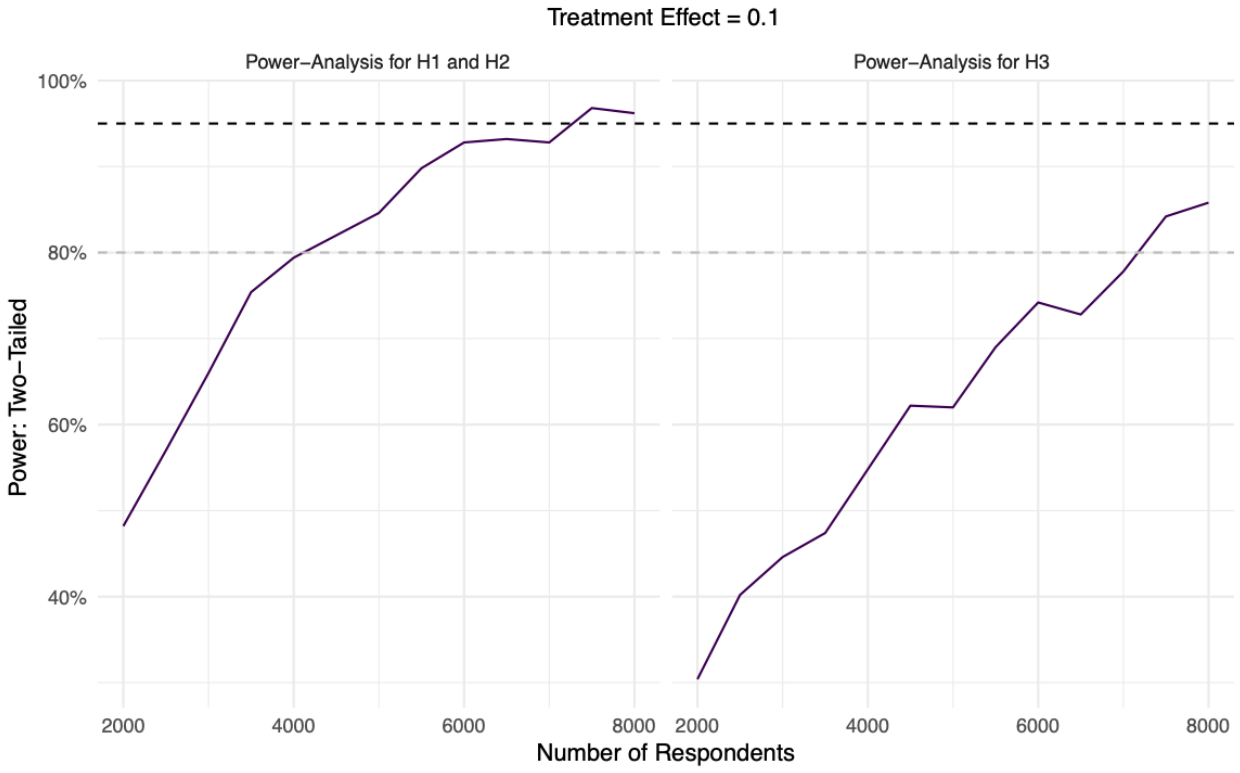


Figure 3: Power Analysis - Pooled



an 11-point scale from `very poorly` (value of `0`) till `very well` (value of 10). See Table 1 `PT-3` for the exact phrasing of the question.

4. *Career perspective* This measure evaluates the politician more broadly by asking whether the politician is likely to move up in leadership position. Respondents answer this question on an 11 point scale, ranging from `very unlikely` (value of `0`) till `very likely` (value of `10`). See Table 1 `PT-4` for the exact phrasing of the question.

Table 1: Survey Questions - DV

Variable	Wording ENG	Wording NL
leadership_eval (PT-1)	To what extent does the following statement apply to [NAME]? [NAME] ...	In hoeverre is onderstaande stelling van toepassing op [NAAM]? [NAAM]...
leadership_eval (PT-1_1)	... has integrity	... is integer
leadership_eval (PT-1_2)	... is competent	... is competent
leadership_eval (PT-1_3)	... is decisive	... is daadkrachtig
leadership_eval (PT-1_4)	... is empathetic	... is empathisch
leadership_eval (PT-1_5)	... is sympathetic	... is sympatiek
leadership_eval (PT-1_6)	... knows what goes on in society	... weet goed wat er leeft in de samenleving
favorable (PT-2)	How disapproving or approving are you of the decision by [NAME] to [NOT] strike a compromise on the [ISSUE]?	Hoe afkeurend of goedkeurend staat u tegenover de beslissing van [NAAM] om [(G)EEN] compromis te sluiten op [ISSUE]?
representation (PT-3)	How well does the following statement describe the politician in question: [NAME] is a good representative of [HIS/HER] constituency.	Hoe goed beschrijft de volgende stelling de desbetreffende politicus: [NAAM] is een goede vertegenwoordiger van [ZIJN/HAAR] achterban.
career (PT-4)	Do you think [NAME] is likely to move up the ladder, for example as a minister in government or a leadership position in parliament?	Denkt u dat [NAAM] waarschijnljk hogerop zal komen, bijvoorbeeld als minister in de regering of een leidinggevende positie in het parlement?

3.2 Control Variables

As control variables, the following *demographics* are measured: gender, age, education, geographical region, level of urbanness, migration background, vote choice in the 2021 parliamentary elections, employment, and income. For the analysis, only variables that are unbalanced over the experimental conditions will be included. Table 2 gives an overview of the questions asked in the survey as well as their English translations.

- *Gender* is measured as `sex` based on the advice of Kieskompas. The answer categories are `Male` (value of `1`), `Female` (value of `0`), and `Other` (value of `999`).
- *Age* is measured as a respondents birth year.
- *Education* is measured as the highest successfully completed level of education, recoded into three categories: High (Dutch levels *HBO* and *WO*), Middle (Dutch levels *MBO*, *HAVO*, and *VWO*), and Low (Dutch levels *LBO*, *MAVO*, *elementary school*, or *none*). I create dummy variables for each level of education with the lowest category as base category.

- *Geographical region* is measured using the *Nielsen districs*, dividing the Netherlands into 1) the 3 major cities plus suburbs, Amsterdam (plus Diemen, Ouder-Amstel, Landsmeer, Amstelveen), Rotterdam (plus Schiedam, Capelle aan den IJssel, Krimpen aan den IJssel, Nederlek, Ridderkerk, Barendrecht, Albrandswaard) and The Hague (plus Leidschendam, Voorburg, Rijswijk, Wassenaar, Wateringen); 2) West (Noord-Holland, Zuid-Holland and Utrecht (excluding the major cities and their suburbs; 3) North (Groningen, Friesland and Drenthe), 4) East (Overijssel, Gelderland and Flevoland); and South (Zeeland, Noord-Brabant and Limburg).
- *Urbanness* Respondents are asked for their postal codes. Using the Dutch Bureau of Statistics's classification we calculate the level of urbanism, ranging from `not at all urban` to `very urban`.
- *Migration background* I use the Dutch Bureau of Statistics's classification for the type of migration background. The Bureau of Statistics distinguishes between `No migration background`, `1st generation Western`, `2nd generation Western`, `1st generation Non-Western`, and `2nd generation Non-Western`.
  I create dummy variables for each category with the `2nd generation Non-Western` as base category.
- *Vote recall* Respondents were asked which party they voted for in the 2021 parliamentary elections. The options were 1) all parties that were elected into parliament –`CDA`, `ChristenUnie`, `D66`, `Denk`, `Forum for Democracy`, `GroenLinks`, `PvdA`, `Animal Rights Party`, `PVV`, `SGP`, `SP`, `VVD`, `50Plus Party` – 2) another party; 3) blanco vote; and 4) a `Don't know` option.
- *Employment* Respondents were asked which category of employment – `Full-time employed`, `Part-time employed`, `Entrepreneur`, `Unemployed and searching for a job`, `Unemployed and not searching for a job or incapacitated`, `Housewife/Househusband or else`, `Retired`, `Student or full-time education` – applied most to them.
- *Income* Respondents were questioned on their monthly income in bins of €500 – `€500 or less`, `€501-€1000`, `€1001-€1500`, `€1501-€2000`, `€2001-€2500`, `€2501-€3000`, `€3001-€3500`, `€3501-€4000`, `€4501-€7500`, `€7501 or more` – as well as giving them the options of `won't say` and `don't know`.

Table 2: Survey Questions - Demographics

Variable	Wording ENG	Wording NL
gender (F1)	What is your gender?	Wat is uw geslacht?
age (F2)	What is your year of birth?	Wat is uw geboortejaar?
education (F9)	What is your highest level of education?	Wat is uw hoogst GENOTEN opleiding
region (F3)	In which region do you live?	In welke regio woont u?
postal_code (F4)	We ask your postal code to be able to make a classification on urbanity (very urban, strongly urban, moderately urban, little urban, not urban). By entering your postal code, you explicitly agree that your region of residence will be used for analysis purposes.	Wij vragen uw postcode om een indeling op stedelijkheid (zeer sterk stedelijk, sterk stedelijk, matig stedelijk, weinig stedelijk, niet stedelijk) te kunnen maken. Door het invoeren van uw postcode gaat u er expliciet mee akkoord dat uw woonregio voor analyse doeleinden wordt gebruikt.
migration_backgr (F5)	Where is your country of birth, Where is your father's country of birth?, and Where is your monther's country of birth?	Wat is uw geboorteland, Wat is het geboorteland van uw vader, en Wat is het geboorteland van uw moeder?
vote_recall (F6)	Which party did you vote for during the last parliamentary elections of March 2021?	Op welke politieke partij heeft u bij de vorige Tweede Kamerverkiezingen van maart 2021 gestemd?
job (F7)	What applies most to you? Are you...	Wat is het meest op u van toepassing? Bent u...
income (F8)	Can you indicate which income groups your net monthly household income falls into?	Kunt u aangeven in welke inkomensgroepen uw netto maandelijks huishoudinkomen valt?

In addition, pre-treatment, respondents' ideological position, position on the issues migration, defense, educaton, and COVID-19 policies, attitudes towards women in politics, and political efficacy are measured in a random order (see Tables 3 and 4). Those variables will only be included in the analyses if balance checks indicate they are necessary. Moreover, the variables will be used to explore heterogeneous relationships.

- *Ideological position* is measured using an 11-point scale ranging from left (`0`) to right (`10`).
- *Position on immigration* is measured using two items. In the first item, participants are asked to estimate the percentage of immigrants living in the Netherlands, which gives an indication of how prejudiced people are towards people with an immigrant background. In the second item, participants are asked to position themselves on whether people with an immigrant background should fully adapt to the Dutch culture on an 11-point scale, ranging from `fully maintain their own values` (value of `0`) to `fully adapt` (value of `10`). Hence, higher values indicate a stronger anti-migration position.
- *Position on defense* is measured by asking people whether or not more money should be spend on defense to keep up with the NATO requirements using a 5-point Likert-scale (`fully disagree`, `disagree`, `neutral`, `agree`, `fully agree`).
- *Position on education* is measured by asking people whether or not more money should be spend on education to maintain the quality thereof using a 5-point Likert-scale (`fully disagree`, `disagree`, `neutral`, `agree`, `fully agree`).
- *Position on COVID-19 policies* is measured by asking people for your opinion on the measures the government chooses to counter the coronavirus, where people can choose for the options: `Much stricter measures to combat the virus (i.e. area ban).`, `Slightly stricter measures to prevent viruses.`, `The measures chosen by the government.`, `Slightly more flexible measures so that only the most vulnerable elderly are protected.`, `Much smoother measures to get the economy and nightlife going immediately.`
- *Attitude towards women in politics* is measured by using the sexism scale developed by Aaldering and Van der Pas. Using five statements (see Table 3) people are asked on a 5-point Likert-scale whehter they `fully disagree`, `disagree`, `neutral`, `agree`, or `fully agree`. The items will be constructed into an additive scale.
- *Political efficacy* is measured using seven statements (see Table 4) a 5-point Likert-scale (`fully disagree`, `disagree`, `neutral`, `agree`, `fully agree`).

3.3 Attention Checks

I include two attention checks in the survey. The first one is after the pre-treatment covariates, the second one is asked just before respondents enter the third round of the experimental treatments. The attention checks are taken from Berinsky, Margolis, and Sances (2014) and adapted to the Dutch context (see also PAPER IR). If a respondents fails the first attention check, a warning appears and the respondent can only continue with the survey once the respondent has correctly answered the question correctly. The second attention check also has a warning – meaning that respondents have to select two options – but if they fail to correctly pass the check, they are excluded. Each excluded respondent due to failing an attention check is replaced with another "good respondent".

Attention Check 1 When a big news story breaks people often go online to get up-to-the-minute details on what is going on. We want to know which websites people trust to get this information. We also want to know if people are paying attention to the question. To show that you have read this much, please ignore the question and select Volkskrant and Metro as your two answers. When there is a big news story, which is the one news website you would visit first? (Please only choose one). Eight (Dutch) news outlets are provided to choose from. Respondents pass the attention check if they select *de Volkskrant* and *Metro*.

Attention Check 2: We would like to get a sense of your general preferences. Most modern theories of decision making recognize that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. To demonstrate that you've read this much, just go ahead and select both red and green among the alternatives below, no matter what your

Table 3: Survey Questions - PreTreatment Questions (1)

Variable	Wording ENG	Wording NL
immigration-percentage (PreT1)	What do you think is the share of immigrants of the Dutch population? By immigrants we mean people born in another country (first generation) and their children (second generation). Enter a percentage between 0 and 100 below, where 0 means that there are no immigrants in the Netherlands at all, 100 means that all residents of the Netherlands are immigrants, and 50 percent means that half of the Dutch population is immigrants.	Wat denkt u dat het aandeel is van immigranten van de Nederlandse bevolking? Onder immigranten verstaan we mensen die zijn geboren in een ander land (eerste generatie) en hun kinderen (tweede generatie). Vul hieronder een percentage in tussen 0 en 100, waarbij 0 betekent dat er helemaal geen immigranten in Nederland zijn, 100 betekent dat alle inwoners van Nederland immigrant zijn, en 50 procent betekent dat de helft van de Nederlandse bevolking immigrant is.
immigration-assimilation (PreT2)	In the Netherlands, some think that Dutch people with a migration background should be able to live here while maintaining their own culture. Others think they should adapt completely to the Dutch culture. Where would you place yourself on a line from 0 to 10, where 0 means preservation of own culture for immigrants and 10 that they should adapt completely?	In Nederland vinden sommigen dat Nederlanders met een migratieachtergrond hier moeten kunnen leven met behoud van de eigen cultuur. Anderen vinden dat zij zich geheel moeten aanpassen aan de Nederlandse cultuur. Waar zou u uzelf plaatsen op een lijn van 0 tot en met 10, waarbij de 0 behoud van eigen cultuur voor allochtonen betekent en de 10 dat zij zich geheel moeten aanpassen?
defense (PreT3)	The Netherlands must increase defense spending to meet NATO requirements.	Nederland moet de uitgaven aan defensie verhogen om zo te voldoen aan de eisen van de NAVO.
education (PreT4)	The Netherlands must increase spending on education to ensure its quality.	Nederland moet de uitgaven aan het onderwijs verhogen om zo de kwaliteit ervan te kunnen waarborgen.
sexism-scale (PreT5)	To what extent do you agree with the following statements:	In hoeverre bent u het eens met de volgende uitspraken:
sexism-scale (PreT5-1)	I hope that in the next ten years the Netherlands will have a female prime minister	Ik hoop dat Nederland in de komende tien jaar een vrouwelijke minister-president krijgt
sexism-scale (PreT5-2)	Men are better at making political decisions than women.	Mannen zijn beter in staat om politieke beslissingen te nemen dan vrouwen.
sexism-scale (PreT5-3)	It is easier in the Netherlands for a man to become a minister than for a woman.	Het is in Nederland voor een man makkelijker om minister te worden dan voor een vrouw.
sexism-scale (PreT5-4)	Political parties must do more to ensure that the number of women in the House of Representatives increases.	Politieke partijen moeten meer doen om te zorgen dat het aantal vrouwen in de Tweede Kamer toeneemt.
sexism-scale (PreT5-5)	Women have fewer opportunities in politics in the Netherlands than men.	Vrouwen krijgen in Nederland minder kansen in de politiek dan mannen.
rile-self (PreT6)	In politics there is often talk about left and right. Where would you put yourself on this scale?	In de politiek wordt vaak gesproken over links en rechts. Waar zou u uzelf op deze schaal plaatsen?

favourite color is. Yes, ignore the question below and select both of those options. What is your favourite color? Six colors are provided to choose from, respondents pass the attention check if they select red and green.

Table 4: Survey Questions - PreTreatment Questions (2)

Variable	Wording ENG	Wording NL
poleff (PreT7)	You will now get to see a number of theses about Dutch politics. To what extent do you agree or disagree with these statements?	In welke mate bent u het eens of oneens met deze stellingen?
poleff (PreT7-1)	The political parties are only interested in my vote and not in my opinion.	De politieke partijen zijn alleen maar geïnteresseerd in mijn stem en niet in mijn mening.
poleff (PreT7-2)	Against their better judgment, politicians promise more than they can deliver.	Tegen beter weten in beloven politici meer dan ze kunnen waarmaken.
poleff (PreT7-3)	Most politicians are skilled people who know what they are doing.	De meeste politici zijn bekwame mensen die weten wat ze doen.
poleff (PreT7-4)	Politicians are able to solve the most important problems.	Politici zijn in staat om de belangrijkste problemen op te lossen.
poleff (PreT7-5)	I feel well represented by Dutch politicians.	Ik voel mij goed vertegenwoordigd door Nederlandse politici.
poleff (PreT7-6)	What in politics is called 'making compromises' is actually just betraying your principles.	Wat men in de politiek 'het sluiten van compromissen' noemt, is eigenlijk gewoon het verraden van je principes.
poleff (PreT7-7)	A strong head of government is good for the Netherlands, even if he stretches the rules a bit to get things done.	Een sterke regeringsleider is goed voor Nederland, ook als deze de regels wat oprekt om dingen voor elkaar te krijgen.
corona-policy (PreT8)	We would like to ask you for your opinion on the measures the government chooses to counter the coronavirus. All in all, are you in favor of:	Hoe denkt u over de maatregelen die de regering kiest om het coronavirus tegen te gaan? Bent u al met al een voorstander van:

3.4 Exclusion Criteria

Participants are required to respond to each question. Participants who fail the second attention check will be excluded but replaced by another participant.

4 Analysis

I test the hypotheses formulated in Section 1 by fitting linear multivariate regressions separately for the four issues. In each model, will estimate the coefficient for gender of the politician (H1), migration background of the politician (H2), and the interaction of gendere and migration background (H3). I will apply a multiverse approach (Steegen et al. 2016; Simonsohn, Simmons, and Nelson 2019). Instead of running various models using the different dependent variables, I will estimate whether the treatment explains the different operationalizations of the dependent variable (see Section 3.1) in one go and present the estimates and confidence intervals for the several combinations. I will only add control variables in the analyses that are unbalanced, as explained in Section 4.1.

4.1 Balance Checks

I will conduct a balance test based on demographics (age, gender, education, geographical region, level of urbanness, migration background, employment, and income), vote choice in the 2021 parliamentary elections, ideological self-placement, political efficacy, attitudes towards women in politics, and positions on the four issues, using the `cobalt` R package (Greifer 2021). If the groups are unbalanced on one of these variables – i.e. standardized mean differences `<0.05` – I will add the covariates to the analyses. I will use the code below to conduct the balance tests (see here for the R script).

```r
covs <- d %>%
  mutate(treatment = paste(name, compromis, sep = "-")) %>%
  select(treatment, PreT1:PreT8, F1:F9)

balanced <-bal.tab(Treatment ~ PreT1 + PreT2 + PreT3 + PreT4 +
                     PreT5 + PreT6 + PreT7 + PreT8 +
                     factor(F1) + F2 + factor(F9) + factor(F3) +
                     factor(F4) + factor(F5) + factor(F6) + F7 + F8,
                     Age + Education + factor(Sex_str), data = covs,
          thresholds = c(m = 0.05))[[1]]

balanced <- balanced %>%
  mutate(variable = c("Immigration (%)", "Immigration", "Defense", "Education", "Sexism",
                      "Ideology", "Poltical Efficacy", "COVID-19",
                      "Gender: Male", "Age", "Education: Middle", "Education: High",
                      "Region: West", "Region: North", "Region: East", "Region: South",
                      "Little Urban", "Moderatly Urban", "Highly Urban",
                      "Very Highly Urban", "Migration Background", "Bij1","CDA",
                      "ChristenUnie", "D66", "Denk", "Forum for Democracy",
                      "GroenLinks", "JA21", "PvdA", "Animal Rights Party", "PVV",
                      "SGP", "SP", "VOLT","VVD", "50Plus Party", "Employment",
                      "Income"),
         variable = factor(variable,
                           levels = c("Immigration (%)", "Immigration", "Defense",
                                      "Education", "Sexism", "Ideology",
                                      "Poltical Efficacy","COVID-19",
                                      "Gender: Male", "Age", "Education: Middle",
                                      "Education: High", "Region: West", "Region: North",
                                      "Region: East", "Region: South",
                                      "Little Urban", "Moderatly Urban", "Highly Urban",
                                      "Very Highly Urban", "Migration Background",
                                      "Bij1","CDA", "ChristenUnie", "D66", "Denk",
                                      "Forum for Democracy", "GroenLinks", "JA21",
                                      "PvdA", "Animal Rights Party", "PVV",
                                      "SGP", "SP", "VOLT","VVD", "50Plus Party",
                                      "Employment", "Income")),
         difference = Diff.Un,2) %>%
  select(variable, difference) %>%
  ggplot(aes(x = variable, y = difference)) +
  geom_point(size = 3, colour = "gray55") +
  theme_bw() +
  labs(x="", y= "Standardized Mean Differences") +
  ggtitle("Covariate Balance") +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_hline(yintercept = 0) +
  geom_hline(yintercept = 0.05, linetype = "dashed") +
  geom_hline(yintercept = -0.05, linetype = "dashed") +
  coord_flip()
```

4.2 Hypothesis 1

I test the *gender hypothesis* using Equation 1 and the code chunk below. The coefficient of β_1 denotes the difference in the evaluation of female and male politicians. The coefficient of β_2 denotes the difference in

the evaluation of politicians who do and do not strike a comprmise. The coefficient of β_3 is the interaction of the β_1 and β_2. If the Average Marginal Effect of β_1 is negative and statistically siginificant when $\beta_2 = 1$ (i.e. politician stroke a compromise), the *gender hypothesis* will be confirmed.

$$\hat{Y} = \beta_0 + \beta_1 Female + \beta_2 Compromise + \beta_3 Compromise * Female + \varepsilon \quad (1)$$

```
df <- d %>% select(PT1:PT4, gender, compromise)

depVarList <- setdiff(colnames(df), c("gender", "compromise"))
allModels <- lapply(depVarList, function(x){
  lm(formula= paste0("`", x, "` ~ gender * compromise"),
     data= d, na.action = na.omit)})

for(i in 1:length(depVarList)){
  if(i==1){
    m <- summary(margins(allModels[[i]]), at = list(compromise = 1)) %>%
      filter(factor=="gender") %>%
      select(AME, lower, upper) %>%
      mutate(y = depVarList[i])
  }
  else{
    tmp <- summary(margins(allModels[[i]]), at = list(compromise = 1)) %>%
      filter(factor=="gender") %>%
      select(AME, lower, upper) %>%
      mutate(y = depVarList[i])
    m <- m %>%
      add_case(tmp)
  }
}
```

4.3 Hypothesis 2

I test the *migration hypothesis* using Equation 2 and the code chunk below. The coefficient of β_1 denotes the difference in the evaluation of politicians with and without a migration background. If β_1 is negative and statistically siginificant, the *migration hypothesis* will be confirmed. The coefficient of β_2 denotes the difference in the evaluation of politicians who do and do not strike a comprmise. The coefficient of β_3 is the interaction of the β_1 and β_2. If the Average Marginal Effect of β_1 is negative and statistically siginificant when $\beta_2 = 1$ (i.e. politician stroke a compromise), the *migration hypothesis* will be confirmed.

$$\hat{Y} = \beta_0 + \beta_1 MigrationBackground + \beta_2 Compromise + \beta_3 Compromise * MigrationBackground + \varepsilon \quad (2)$$

```
df <- d %>%
  mutate(migration = recode(name,
                            `Rachid Amezian` = 1,
                            `Rachida Amezian` = 1,
                            .default = 0)) %>%
  select(PT1:PT4, migration, compromise)

depVarList <- setdiff(colnames(df), c("migration", "compromise"))
allModels <- lapply(depVarList, function(x){
  lm(formula= paste0("`", x, "` ~ migration * compromise"),
     data= d, na.action = na.omit)})
```

```
for(i in 1:length(depVarList)){
  if(i==1){
    m <- summary(margins(allModels[[i]]), at = list(compromise = 1)) %>%
      filter(factor=="migration") %>%
      select(AME, lower, upper) %>%
      mutate(y = depVarList[i])
  }
  else{
    tmp <- summary(margins(allModels[[i]]), at = list(compromise = 1)) %>%
      filter(factor=="migration") %>%
      select(AME, lower, upper) %>%
      mutate(y = depVarList[i])
    m <- m %>%
      add_case(tmp)
  }
}
```

4.4 Hypothesis 3

I test the *intersection hypothesis* using Equation 3 and the code chunk below. The coefficient of β_1 denotes the difference in the evaluation of female politicians with a migration background and the other politicians. If β_1 is negative and statistically siginificant, the *gender hypothesis* will be confirmed. The coefficient of β_2 denotes the difference in the evaluation of politicians who do and do not strike a comprmise. The coefficient of β_3 is the interaction of the β_1 and β_2. If the Average Marginal Effect of β_1 is negative and statistically siginificant when $\beta_2 = 1$ (i.e. politician stroke a compromise), the *intersection hypothesis* will be confirmed.

$$\hat{Y} = \beta_0 + \beta_1 Intersection + \beta_2 Compromise + \beta_3 Compromise * Intersection + \varepsilon \tag{3}$$

```
df <- d %>%
  mutate(intersection = recode(name,
                          `Rachida Amezian` = 1,
                          .default = 0)) %>%
  select(PT1:PT4, intersection, compromise)

depVarList <- setdiff(colnames(df), c("intersection", "compromise"))
allModels <- lapply(depVarList, function(x){
  lm(formula= paste0("`", x, "` ~ intersection * compromise"),
     data= d, na.action = na.omit)})

for(i in 1:length(depVarList)){
  if(i==1){
    m <- summary(margins(allModels[[i]]), at = list(compromise = 1)) %>%
      filter(factor=="intersection") %>%
      select(AME, lower, upper) %>%
      mutate(y = depVarList[i])
  }
  else{
    tmp <- summary(margins(allModels[[i]]), at = list(compromise = 1)) %>%
      filter(factor=="intersection") %>%
      select(AME, lower, upper) %>%
      mutate(y = depVarList[i])
    m <- m %>%
      add_case(tmp)
```

```
  }
}
```

4.5 Pooled Data

Using the pooled data we will estimate a within groups fixed effects model. $Y\hat{Y}_{r,i,t}$ in Equation 4 denotes the evaluation of a politician of respondent r, during issue i and at experimental round t – ranging from round 1 to round 4. α_i is the issue specific intercept, and γ_t is the experimental round specific intercept.
The standard errors are clustered at the individual level.
If the Average Marginal Effect of β_1 is negative and statistically siginificant when $\beta_2 = 1$ (i.e. politician stroke a compromise), the *intersection hypothesis* will be confirmed.

$$\hat{Y}_{r,i,t} = \beta_0 + \beta_1 Intersection_{r,i,t} + \beta_2 Compromise + \beta_3 Compromise * Intersection + \alpha_i + \gamma_t + \varepsilon_{r,i,t} \quad (4)$$

4.6 Statistical Significance

All the hypotheses are directional, and therefore all of the tests will be one-tailed. I will use an α-value of `0.05` as the value for statistical significance in all models above.

5 Stimulus Material

Table 5: Experimental Conditions

	Name Politician	Gender	Migration Background	Compromise
1	Rachid Amezian	Men	Yes	Yes
2	Rachid Amezian	Men	Yes	No
3	Karel van der Kleijn	Men	No	Yes
4	Karel van der Kleijn	Men	No	No
5	Rachida Amezian	Women	Yes	Yes
6	Rachida Amezian	Women	Yes	No
7	Karin van der Kleijn	Women	No	Yes
8	Karin van der Kleijn	Women	No	No

Table 5 show the manipulations of the `2*2*2` experimental design, leading to 8 experimental groups. All stimulus material is developed by the Network Institute of the *Vrije Universiteit Amsterdam* and can be accessed here. The treatment's texts are in Dutch. The English translation is displayed in Table 6.

Table 6: Text of Experimental Treatment

Title	Politician makes [**no/ **] compromise on [**expansion creative elementary school activities/ intensifying defense cooperation/ migration freeze in Western countries/ Corona emergency legislation**]
First Paragraph	In this term, parliament faces a crucial crossroads regarding an [**expansion of creative elementary school activities/ intensification of defense cooperation/ migration freeze in Western countries/ extension of Corona emergency legislation**]. With the deadline for a vote fast approaching, parliament must determine the future of the policy on [**expanding creative elementary school activities/ intensifying defense cooperation/ migration freeze in Western countries/ Corona emergency legislation**] in the Netherlands. A number of non-partisan organizations have urged parliament to reach a compromise on this issue. Time is running out.
Second Paragraph	For this important bill to pass, politicians from different political parties will have to put aside their differences and reach a compromise. While many in parliament are still debating their positions, some members of both left and right parties have already shown a willingness to compromise. One of those who [**won't/ will**] compromise is the politician [**Rachid Amezian/ Rachida Amezian/ Karin van der Kleijn/ Karel van der Kleijn**]. [**Rachid Amezian/ Rachida Amezian/ Karin van der Kleijn/ Karel van der Kleijn**] has already publicly stated that [**he / she**] wants to make [**no/ a**] compromises and will vote [**against/ for**] this bill. Although the vote is still a week away, [**Rachid Amezian/ Rachida Amezian/ Karin van der Kleijn/ Karel van der Kleijn**] has told numerous media sources that [**he/ she**] will not change [**his/ her**] vote. [**'This is just not an issue on which I want to compromise'/ 'This is an issue on which I am willing to compromise'**] , said [**Rachid Amezian/ Rachida Amezian/ Karin van der Kleijn/ Karel van der Kleijn**]. 'My voters knew this when they elected me.'

References

Aaldering, Loes, and Rens Vliegenthart. 2016. "Political Leaders and the Media. Can We Measure Political Leadership Images in Newspapers Using Computer-Assisted Content Analysis?" *Quality & Quantity* 50 (5). Springer: 1871–1905. https://doi.org/10.1007/s11135-015-0242-9.

Bauer, Nichole M, Laurel Harbridge Yong, and Yanna Krupnikov. 2017. "Who Is Punished? Conditions Affecting Voter Evaluations of Legislators Who Do Not Compromise." *Political Behavior* 39 (2). Springer: 279–300. https://doi.org/10.1007/s11109-016-9356-6.

Berinsky, Adam J, Michele F Margolis, and Michael W Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58 (3). Wiley Online Library: 739–53. https://doi.org/10.1111/ajps.12081.

Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. 2019. "Declaring and Diagnosing Research Designs." *American Political Science Review* 113 (3). Cambridge University Press: 838–59. https://doi.org/10.1017/S0003055419000194.

Doherty, David. 2013. "To Whom Do People Think Representatives Should Respond: Their District or the Country?" *Public Opinion Quarterly* 77 (1). Oxford University Press UK: 237–55. https://doi.org/10.1093/poq/nfs052.

Eagly, Alice H. 2007. "Female Leadership Advantage and Disadvantage: Resolving the Contradictions." *Psychology of Women Quarterly* 31 (1). SAGE Publications Sage CA: Los Angeles, CA: 1–12. https://doi.org/10.1111/j.1471-6402.2007.00326.x.

Elis, Roy, D Sunshine Hillygus, and Norman Nie. 2010. "The Dynamics of Candidate Evaluations and Vote Choice in 2008: Looking to the Past or Future?" *Electoral Studies* 29 (4). Elsevier: 582–93. https:

//doi.org/10.1016/j.electstud.2010.04.005.

Greifer, Noah. 2021. "Cobalt: Covariate Balance Tables and Plots. R Package Version 4.3.1." https://cran.r-project.org/web/packages/cobalt/index.html.

Simonsohn, Uri, Joseph P Simmons, and Leif D Nelson. 2019. "Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications." *Available at SSRN 2694998*. https://doi.org/10.2139/ssrn.2694998.

Steegen, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. "Increasing Transparency Through a Multiverse Analysis." *Perspectives on Psychological Science* 11 (5). Sage Publications Sage CA: Los Angeles, CA: 702–12. https://doi.org/10.1177/1745691616658637.