The influence of different voting systems on polarisation

Reducing voting-induced polarisation through multi-votes systems

Master Thesis Economics and Business
Behavioural Economics
24 July 2017

Abstract:
Polarisation is a worldwide phenomenon which frequently dominates the news. This research's contribution to the literature about voting-induced polarisation is two-fold. First, it uses a new methodology to investigate the influence of the act of voting on polarisation. Existing literature focuses on historical data while this thesis makes use of an online controlled experiment. Second, this thesis investigates the influence of different voting systems on polarisation. Results show that voting for one candidate increases polarisation within the electorate. Next to this, voting for two candidates increases polarisation less than voting for one candidate. Despite the fact that the latter result is not significant, it serves as an important starting point for future research focusing on ways to decrease polarisation.

Author: Rik Bross
Student ID number: 360566
Supervisor: Dr. Georg D. Granic
Second assessor: Dr. Paul van Bruggen

Acknowledgements

I would like to thank my supervisor, Georg Granić, for his inspiration regarding this topic and his guidance during this thesis. Next, I would like to thank Georg Granić and Paul van Bruggen for their flexibility to schedule my defence before September to ensure that I can start working from 1 September onwards. Furthermore, I would like to thank my family who has supported me (mentally and financially) during my study. Lastly, I would like to thank my girlfriend for her support and help to improve my English writing skills.

Table of Content

List of Figures and Tables

1. Introduction

Polarisation is a worldwide phenomenon which frequently dominates the news. Polarisation can be described as a sharp division into two different groups or sets of opinion or beliefs. In the past years, multiple examples show the massive impact of polarisation, both worldwide and on a country level. The effects of polarisation can be tremendous in economic, social and political terms. One recent example is the Brexit referendum in June 2016 (Saltzman, 2016). Estimations show that the economic consequences of Brexit will result in a 3% lower GDP for Great-Britain in 2020, compared to a situation in which Great-Britain did not leave the EU. These costs equal a cost of approximately 2200 GBP per household. Furthermore, in the long run, structural impacts through channels of capital, immigration, and lower technical progress are expected in Great-Britain due to the Brexit (Kierzenkowski et al., 2016). Also, the referendum in Turkey in 2017 (Saleem, 2017), the election of Trump (Heaney, 2017), and the increasing nationalism in European countries (Powel et. al., 2017) are recent examples of the economic, social, and political impact of polarisation.

The previous examples show that polarisation is evident at this moment. However, polarisation has been of interest to many researchers for decades. For example, Wilks (1975) studied the polarised Asante in Africa during the 19th century. During the years many different research directions in many various fields were explored to find factors causing polarisation. Recently, a new research direction is added to the possible influences of polarisation. Mullainathan & Washington (2009) provided evidence that voting itself may cause polarisation. They applied the theory of cognitive dissonance (Festinger, 1957) to political polarisation. This theory supports that choices result in attitude change after the choice is made. Applying this to a political field, this means that the act of voting for a party or candidate could lead to an increasingly polarised electorate.

Built on the research of Mullainathan & Washington (2009), various researchers investigated the influence of the act of voting on polarisation. Different data from numerous countries resulted in a wide variety of conclusions. The contribution of this thesis to the literature is twofold. First, The literature has typically overlooked that institutions, for example, electoral systems, may moderate or interact with voting-induced polarisation. The aim of this thesis is to bridge this gap by investigating whether a voting system in which people vote for more than one candidate influences the level of polarisation less than a voting system in which people vote for one candidate. Second, this thesis provides a new methodology for investigating voting-induced attitude change. Existing research is based on historical data while this thesis aims for an online controlled economic experiment. This new method is needed to investigate the influence of different voting systems on polarisation.

For this research, an experiment is designed to examine a causal relationship among different voting systems and polarisation. In total, 139 subjects are divided randomly into three separate groups. One control group in which people rated the candidates two times before voting for one candidate and two treatment groups in which subjects either vote for one or two candidates before they rated the candidates again. On average, voting for one candidate increases the difference in rating scores between the chosen candidate and the unchosen candidates. This means that a voting system in which people vote for one candidate increases polarisation. On average, voting for two candidates increases the difference in rating scores between the chosen candidates and the unchosen candidates less than voting for one candidate. However, this result is not significant. This means that one cannot conclude at a 5% significance level that a voting system in which people vote for two candidates increases polarisation less compared to a one-candidate voting system. Nevertheless, the obtained results indicate that there could be a significant effect which is capturable by improving the experimental design. Therefore, more research on the effects of alternative voting systems is needed.

The remainder of this thesis is structured as follows. Section 2 starts with a literature review which both the terms polarisation and cognitive dissonance explains in more details. After, the both hypotheses are described and designed. Section 3 gives an overview of the experimental design and the obtained data. Section 4 provides an overview of all the analyses and results. This section includes one-sample tests, multiple samples tests, and regression analyses to conclude on both hypotheses. Section 5 starts with a discussion and ends with the conclusion of this research.

2. Literature Review

This research builds up on existing literature by reviewing the influence of different voting systems on polarisation. Before constructing the both hypotheses, it is important to review the existing literature regarding polarisation and cognitive dissonance. This section presents and explains the existing literature. The Sections 2.1 and 2.2 focus on polarisation and cognitive dissonance respectively, and Sections 2.3 formulates the two hypotheses of this research.

2.1 Polarisation

Esteban & Ray (1994) described polarisation as follows: Consider a particular distribution on Y, which can be anything. A population can then be divided into several groups that have a large degree of within-group homogeneity, while the comparison between groups displays a considerable degree of heterogeneity. To be more concrete, when groups are polarised, the individuals of a specific group are looking for strong similarities within the group and on the other hand for strong differences with other groups. Polarisation can occur in several ways and can influence individuals and even entire nations. An excellent example of creating polarisation comes from a quote George W. Bush Jr. stated after the attacks on 9/11. Bush Jr. (2001): *''Every nation, in every region, now has a decision to make. Either you are with us, or you are with the terrorists''*. With this statement, Bush Jr. aimed for perfect within-group homogeneity and perfect between group heterogeneity.

Polarisation is a phenomenon which is of interest in multiple research fields for many years. Hariharan & Pople (1973), McPhee (1989) and Wilks (1975) are prominent early examples of research on polarisation in different research fields. Political polarisation is a research direction in which recently notable causations are examined. First, this research was mostly focussed on the differences in attitudes between the electorate and people who were not allowed to vote (Crosby & Taylor, 1983; Anderson et al., 2004). Mullainathan & Washington (2009), however, provided evidence that the act of voting itself may cause polarisation within the electorate. A vote for a particular candidate leads to an increasingly favourable attitude towards this candidate. On the other hand, a vote for a particular party or candidate leads to a decreasingly favourable attitude towards the rejected parties or candidates. These attitude changes result in an increase of polarisation on the electorate (Mullainathan & Washington, 2009).

McGregor (2013) was the first who mentioned the importance of a distinction between election systems when comparing the level of polarisation after the act of voting. A system such as the US

where two parties dominate the election creates an 'us' versus 'them' dynamic (Richardson, 1991). On the other hand, when more than two parties are eligible, this dynamic is less clear. Positive attitudes towards a particular party do not necessarily mean negative feelings towards all other parties. This is particularly the case when a party is not the primary opponent of the preferred party (McGregor, 2013). Potential mechanisms which cause voting-induced polarisation are discussed in Section 2.2.

2.2 Cognitive dissonance

Festinger proposed the cognitive dissonance theory in his book 'A Theory of Cognitive Dissonance' (1957). Festinger stated that when related cognitions are inconsistent, a sense of psychological discomfort can arise and that people try to reduce this discomfort by adjusting these cognitions until they are consistent. In other words, people can reduce discomfort caused by inconsistent cognitions by changing these exact cognitions. Cognitions are all processes by which the sensory input is transformed, reduced, elaborated, stored, recovered and used (Chomsky, 1959). Festinger (1957) replaced the word ''inconsistency'' with the term ''dissonance'' and the term ''consistency'' with the term ''consonance'' because he argued these terms to be more neutral and less of a logical connotation. As an example, Festinger (1957) compared cognitive dissonance to hunger which leads to action to reduce hunger. Reducing dissonance is a basic process in humans which can arise in a wide variety of contexts (Festinger, 1957).

The free choice paradigm is one paradigm which results from the cognitive dissonance theory (Cooper, 2007). This paradigm states that people who rate alternatives before and after choosing an alternative, increase their rating score for the chosen alternative and/or decrease their rating score for the unchosen alternative (Brehm, 1956). Recently, Egan et al. (2007) run an experiment where subjects were asked to rate a given series of presents. After, participants could choose one of the presents and then they had to rate the presents again. Egan et al. (2007) found that participants, on average, rated the presents they chose higher after they have been selected than before. Also, subjects rated, on average, the rejected items lower after they chose the other present. This finding is in line with Festinger (1964), who stated that to reduce dissonance, people adjust their impressions of alternatives, by creating positive evaluations for chosen alternatives and negative evaluations for rejected alternatives.

Psychological research also mentions other mechanisms that could produce a similar effect as cognitive dissonance. For example, the self-perception theory of Bem (1967) stated that individuals

infer their opinions from their actions. This self-perception theory seems to be a second explanation for the impact of behaviours on beliefs. Nevertheless, more and more literature provided evidence that dissonance is the unpleasant state of arousal and the self-perception theory is now more seen as an explanation for dissonance instead of another mechanism (Hogg & Cooper, 2003).

The first formulation of Festinger (1956) is based on the idea that people may unconsciously adjust their attitudes to have them fit external facts people cannot change. Festinger & Carlsmith (1959) elaborated on this by investigating how people react to their own behaviour when this behaviour does not confirm their preferences. After conducting a laboratory experiment[1], Festinger & Carlsmith (1959) provided evidence that if persons are induced to do or say something which is inconsistent with their private opinions, there will be a tendency for them to change that view in such a way they bring it into correspondence with what they have done or said. Also, the larger the pressure used to elicit the behaviour, the weaker this tendency will be. Aronson (1992) elaborated by providing evidence regarding the role of behaviour in the reinforcement of prior attitudes, even when the first largely conforms to the second. Attitudes are a person's perceived favourability towards a specific subject (Zanna & Rempel, 1988). Subjects are influenced by cognitive knowledge or beliefs, feeling, and behavioural factors (Worchel et al., 2000). The cognitive dissonance theory can be used to explain the changes in attitudes over time. An important direction of cognitive dissonance research covers the exact impact that behaviour has on attitudes. These studies have also explored the relevance of cognitive dissonance to a large variety of contexts (Aronson, 1999). One of these contexts is the effect of the act of voting on cognitive dissonance.

Elections are a great opportunity to test the cognitive dissonance theory. People are forced to choose between alternatives during an election, and after voting, knowledge of behaviour becomes a cognition. The voting could lead to dissonance and, consequently, a shift in attitude towards a candidate to reduce this dissonance. This attitude shift can happen in three different ways. People could increase the rate of the chosen alternative, decrease the rate of the rejected alternative or even both (McGregor, 2013). Festinger (1964) called this effect the ''spreading of alternatives''. Voting for a party may lead to more favourable cognitions towards the party than before the voting. Even if there is no clear conflict between attitudes and behaviour, there could be a modest dissonance that people want to reduce by changing their attitudes, to make sure that they like their presented behaviour even

[1] Festinger & Carlsmith conducted a laboratory experiment in which subjects were subjected to a boring experience and then paid to tell someone that the experience had been interesting and enjoyable. The amount of money paid the subject was varied. The private opinions of the subjects concerning the experience were then determined.

more clearly (Bølstad et al., 2013). Nevertheless, not all voting leads to dissonance. If people could vote fully according to their attitudes, their behaviour is consonant with their attitudes. However, people often have to make voting decisions that are not in full accordance with their beliefs. People often vote strategically or compromisingly which results in cognitive dissonance. Lastly, individuals who do not vote do not have this process of adjusting their attitudes because they avoided the potential dissonance (Festinger, 1964).

The act of voting, therefore, can result in attitude change. This attitude change can happen through a variety of cognitive, affective, and behavioural factors. These factors also include some necessary drivers. Examples are the level of importance a voter puts on a decision, partisan attachment, unpleasant effort during the process, whether their vote is a winning or losing vote, and the moment in time when making the final decision of the vote (Beasley & Joslyn, 2001 and Mullainathan & Washington, 2009). Models of Aldrich (1993), Grossman & Helpman (2001) and Coate & Conlin (2004) assume that preferences are also a driver in voting decisions. However, multiple experiments have provided evidence that this causation may also run in the opposite direction, so that action themselves drives preferences and beliefs to reduce cognitive dissonance.

2.3 Research question and hypotheses

This thesis combines the literature on political polarisation and cognitive dissonance to investigate the research question: Do different voting systems lead to different levels of voting-induced polarisation? Specifically, a system in which people vote for one candidate is compared to an alternative voting system in which people vote for more than one candidate. In order to answer the research question, two different hypotheses need to be reviewed first.

2.3.1 Hypothesis 1: The influence of voting on polarisation due to cognitive dissonance

Mullainathan and Washington (2009) stated that the endogeneity of a voting decision leaves them unable to treat previous results in the literature as evidence of a causal link between voting and increased polarisation. Mullainathan & Washington (2009) added that voter attendance correlates with voter attitude and that existing literature fails to control for this. Mullainathan & Washington (2009) used age restriction as a variable to establish a causal relationship between the act of voting on polarisation. Comparing the change in polarisation between people who were not allowed to vote (16 and 17 years old) and individuals who were just allowed to vote (18 and 19 years old) Mullainathan & Washington (2009) conclude that voting leads to a higher level of polarisation due to dissonance

reduction. Beasley & Joslyn (2001) and Mullainathan & Washington (2009) used evaluative distance as a measurement to measure the change in polarisation. This is the difference between party evaluation scores for chosen and rejected parties after voting, accounted for the pre-voting differences in evaluation scores.

Bølstad et al. (2013) argued that differentiating between cognitive dissonance and the alternative mechanisms can be done by testing the effect of the act of voting in a setting in which there is no correlation between actual choice and true party preference. Elinder (2012) argued that polarisation effects took place already one month before the elections and concluded that based on his research there is no effect of voting on attitudes through cognitive dissonance. McGregor (2013) on his part, questioned the conclusions of Elinder (2012) because of a needed distinction between two-party and multiple party systems. Section 2.3.2 of this research elaborates on this specific distinction.

Over the years, researchers disagreed and elaborated on each other's research about the effect of voting on polarisation due to cognitive dissonance reduction. Different conclusions are drawn from various results and different methodologies. All performed research has two important things in common. Firstly, all the conclusions are based on historical data, and secondly, all the research is done using voting systems in which people vote for one candidate or party. This thesis contributes to the literature in twofold. First, by investigating the effect of voting on polarisation due to dissonance reduction in a controlled economic experiment and second, by investigating the difference between voting systems on polarisation. In order to study the difference between voting systems, the effect of voting on polarisation needs to be investigated first using a voting system in which people vote for one candidate. For this reason and based on the previous literature the first hypothesis is stated as follows:

Hypothesis 1: A voting system in which people vote for one candidate increases the polarisation of the electorate.

2.3.2 Hypothesis 2: The influence of different voting systems on polarisation

As already mentioned in Section 2.3.1, McGregor (2013) questioned the conclusions of Elinder (2012) because of a needed distinction between two-party and multiple-party systems. In a multi-party system, where party differences are smaller, and the partisan attachment is low, the party ratings are more volatile and choice induced attitude changes are more likely to detect compared to a two-party system (LeDuc et al., 1984; McGregor, 2013). However, not only the party system a country uses could influence the evaluative distance, but also the used election system may affect this effect.

Across the world, there are multiple different election systems. If the electorate votes for a candidate rather than for parties, it is called non-party list system or PR system (Rule, 1987). Countries such as the U.S.A., Britain, Australia, Canada, New Zealand, and France, and the electoral systems of Ireland and Japan make use of such a system (Farrel, 2011). On the other hand, if the electorate votes for a party rather than for a candidate it is called party list system or proportional representation system (Rule 1987). Most of the European countries make use of such a system. The list system is the most used system in the world (Farrel, 2011). One could also make a distinction between proportional and non-proportional systems. Proportional systems ensure that the number of seats each party wins reflects the number of votes a party has received as closely as possible. On the other hand, non-proportional systems ensure that one party has a clear majority of seats (Norris, 2004).

West-Germany was the first country in which an election system is used in which the electorate has more than one vote. The electorate selected half of the seats by using a list/PR system and the other half by using single member districts. Each voter had one vote for the individual candidate in the small district and one for a party in the larger PR district (Lijphart & Grofman, 1984). Nowadays, the structure of German 'Bundestag' elections exists of two votes: one for a constituency candidate and the second for a party list. About 20% of the voters split their ticket (Pappi & Thurner, 2002). Furthermore, the majority of the Spanish Senate and the Belgian city council elections make use of 'limited voting.' Limited voting means that a voter has multiple votes, but less than the available seats (Lijphart & Grofman, 1984). The elections for the Australian House of Representatives and the presidential elections in Ireland make use of an alternative voting system. Here, the electorate has to rank the candidates (Norris, 1997).

The influence of these less common voting systems on polarisation is not yet investigated in isolation and compared with the more common voting systems. Using the same evaluative distance measurement as Beasley & Joslyn (2001) and Mullainathan & Washington (2009) to measure polarisation, one could investigate whether this spread increases less when using an alternative voting system. People may feel less urge to vote strategically or compromisingly when they are allowed to vote for more than one option. For this reason, this thesis tests a system in which the electorate votes two candidates and compare this with a system in which the electorate votes for one candidate. This particular system represents one alternative voting system. Therefore, the second hypothesis states:

Hypothesis 2: A voting system in which people vote for two candidates increases the polarisation of the electorate relatively less than a voting system where people vote for one candidate.

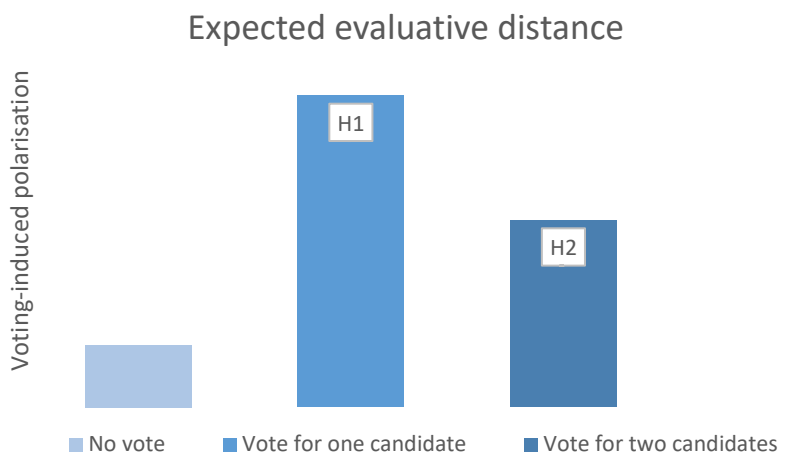

Figure 1: Expected evaluative distance of different treatments

Figure 1 shows a schematic overview of the expected results of the two hypotheses tested in this experiment. The literature provides evidence that the act of voting itself causes polarisation. Due to dissonance reduction, one should expect that people who voted for a candidate or party evaluate this chosen candidate or party higher than the rejected alternatives. However, if people vote for more than one party or candidate, this evaluative distance could be relatively lower because people may vote less strategic or compromising. The evaluative distance between the groups is a proxy for polarisation in this thesis. Section 3 gives a detailed description of the experimental design to test these two hypotheses.

3. Experimental Design & Data

This section introduces the experimental design and presents the data resulting from the experiment. Section 3.1 gives a global overview of the experiment. After this, a detailed description of the candidates (Section 3.2), the rating systems (Section 3.3), the choice tasks (Section 3.4) and the control variables (Section 3.5) are presented to explain the procedure of the experiment in more detail. Section 3.6 presents the obtained data from the performed experiment.

3.1 Global overview experiment

	Control Group	Treatment group 1	Treatment group 2
1	Introduction		
2	Description candidates		
3	Rating candidates for the first time		
4	Filler choice task	Vote-for-1 task	Vote-for-2 task
5	Rating candidates for the second time		
6	Vote-for-1 task	Filler choice task	Filler choice task
7	Political placement		
8	Demographic questions		
9	Closing		

Table 1: Global overview experiment

Table 1 gives a global overview of the procedure of the experiment. The experiment is conducted using the online survey program Qualtrics. All the participants were able to participate anonymously through an online link. This link is distributed through different (personal) channels to reach as many people as possible. The subjects are randomly divided into one of the three distinct groups[2]. However, due to the expected lower variation in the control group where subjects do not vote before the second rating task, the software is programmed in such a way that approximately 20% of the subjects participate in the control group, 40% in treatment group 1 and 40% in treatment group 2. Every group follows the same procedure, consisting of a rating task, followed by a choice task, then again a rating task, and ending with another choice task.

The experiment starts with a brief introduction to the experiment in which the subjects are participating. This introduction explains the rules of the game but gives no information about the goals

[2] Since the academic year 2016-2017 every EUR student has a premium account for Qualtrics which make it possible to use randomizers and other features to conduct an online experiment.

of the experiment to control for experimenter demand effects[3]. Furthermore, no information about other subjects is given to satisfy the privacy precept of Smith (1982). Section 3.1.1 explains the five precepts of Smith (1982) which need to hold for a controlled economic experiment in more detail.

During step two and three, subjects can read the descriptions of the four different candidates and have to rate each candidate. Every candidate makes five statements about what they want to implement after the election. The higher a subject rates a candidate, the more favourable a subject feels towards a candidate. Section 3.2 and Section 3.3 give a more detailed explanation about the candidates and the rating system.

Step four exists of the first choice task for all the participants. The control group starts with a filler task, while both treatment groups start with the voting task. The first treatment group makes use of a vote-for-1 system, which is needed to test the first hypothesis. The second treatment group makes use of a vote-for-2 system, which is necessary to test the second hypothesis. A more detailed description of the different choice tasks is given in Section 3.4.

During step five and six, the subjects have to rate the candidates again and finish the second choice task. The difference in spread between the rating tasks for each group needs to be compared to test the two hypotheses. Based on the literature it is expected that the spread of the first treatment group is relatively larger than the spread of the control group and the second treatment group and that the spread of the second treatment group is relatively larger than the spread of the control group. The second choice task of the control group exists of the voting task, using the same vote-for-1 system as the first treatment group. The second choice task for the both treatment groups exists of the same filler task as the control group performed the first choice task. Again, a more detailed explanation about the candidates and the rating system are given in Section 3.2 and Section 3.3.

Step seven and eight exists of the same questions for each group, which can be used as control variables. More information about these questions is given in Section 3.5. The experiment ends for each subject with the possibility to fill in their e-mail address which gives a one out of 60 chance of winning 20 euro.

[3] Experimenter demand effects are changes in behaviour by subjects due to hints in the information about what constitutes 'appropriate' behaviour. This effect can lead to biased answers, especially when they are positively correlated with the true experimental predictions (Zizzo, 2010).

3.1.1 Controlled economic lab experiment

One of the contributions of this research to the literature is the use of an experiment to investigate the relationship between voting and polarisation due to dissonance reduction. In order to conclude from an experiment some sufficient conditions need to hold, who are known as the five precepts of experimental economics (Smith, 1982). Nonsatiation[4] and saliency[5] both need to hold for an economic experiment and dominance[6] and privacy[7] both need to hold for a controlled economic experiment. Control can be applied by using a reward system and a property right system to generate monetary value on outcomes. Parallelism[8] is the last precept that needs to hold for a controlled economic experiment.

The application of these precepts to this experimental design is needed to ensure control. Nonsatiation is most likely satisfied, where subjects prefer a higher rated candidate over a lower rated candidate. Saliency is harder to satisfy using this experimental design. Incentivising choices is almost impossible in this context and using this experimental design. However, Vossler & Kerkvliet (2003) and Delavande & Manski (2015) provided evidence that political choices in surveys or experiments match actual voting outcomes because elections are mostly considered as very important, even when it is hypothetic. Furthermore, subjects are asked to place the candidates and themselves in a political field to check whether subjects made choices which can be considered as consistent choices. In Section 4.4 these results are investigated to check whether the intrinsic motivation is strong enough to expect saliency satisfied in this case. Also, saliency means that you pay what you promise and do not deceive the subjects. To satisfy this precept, two subjects are randomly picked, and both won 20 euros[9].

Satisfying the dominance principle is a leap of faith in this experiment, meaning it cannot be known for sure whether this principle is truly satisfied. From every 60 participants, one was randomly picked and

[4] **Nonsatiation**: ''Given a costless choice between two alternatives, identical except that the first yields more of a reward medium (for example, U.S. currency) than the second, the first will always be chosen over the second, by an autonomous individual. Hence utility, U(V), is a monotone increasing function of the monetary reward, U'> 0, where V is dollars of currency'' (Smith, 1982).

[5] **Saliency**: ''Individuals are guaranteed the right to claim a reward which is increasing in the goods outcomes, xi, of an experiment and vice versa; individual property rights in messages, and how messages are to be translated into outcomes are defined by the institution of the experiment'' (Smith, 1982).

[6] **Dominance**: ''The reward structure dominates any subjective costs (or values) associated with participation in the activities of an experiment'' (Smith, 1982).

[7] **Privacy**: ''Each subject in an experiment is given information only on his/her own payoff alternatives'' (Smith, 1982).

[8] **Parallelism: ''**Propositions about the behaviour of individuals and the performance of institutions that have been tested in laboratory microeconomies apply also to nonlaboratory microeconomies where similar ceteris paribus conditions hold'' (Smith, 1982).

[9] The supervisor of this thesis, Georg Granic, randomly picked two subjects to ensure an honest way of selecting.

to win 20 euros after finishing the experiment. Most of the subjects are students which make it plausible to argue that the reward structure dominates the personal costs of paying attention for 10 minutes. The privacy principle is most likely to hold in this experiment. As already mentioned, the subjects only receive information to answer the questions and not about other participants or goals of the experiment. Another important matter to satisfy privacy is to make sure subjects can participate anonymously. For this reason, this experiment does not take place in a lab, but through an anonymous online link. Satisfying the parallelism precept is also a leap of faith in this experiment. However, a theory is tested and not the 'real-life behaviour', which makes it plausible to expect that the general laws of behaviour in this experiment apply everywhere.

Concluding, some of the precepts are expected to hold. However, the monetary incentives are very low which means this experiment cannot be considered as a fully controlled economic experiment. Especially on an individual question level; there is no monetary incentive to answer truthfully. Subjects, therefore, need to be intrinsically motivated to answer truthfully about their political positions (Morgan & Stocken, 2008). A closer look at the outcomes of the control questions, therefore, is needed to make sure that the results of this experiment can be used to conclude. The results of these tests are given in Section 4.4.

3.2 The Candidates

The experiment starts with descriptions of the four different candidates who are participating in a hypothetical election. Research of McDermott (1998) shows that the electorate can be influenced by factors such as race, gender, age and appearance of the candidates. For this reason, the four different candidates in this experiment are labelled with neutral colours[10]. The most-left candidate is labelled Turquoise, the middle-left candidate is labelled Beige, the middle-right candidate is labelled Violet, and the most-right candidate is labelled Peach. Next, no further information is given about the candidates to ensure 'clean' statements of the candidates.

Another possible effect to control for is the 'order effect', which can be the result of the order in which the candidates are presented to the subjects. Huber et al. (2009) show that the first candidate in an election increases the change of being chosen with 4 to 17 percentage points. Despite the fact that

[10] Many colours are connected to a political party or stream. For example, red is generally associated with political left wing parties. These colours could therefore influence the subject's evaluation of a candidate. It differs per country and continent which colours are associated with which parties or streams. Therefore, neutral colours are selected, which are nowhere in the world obvious associated with a political party or stream to control for any possible influence of the colours.

this result is only found in the New York Primary election and not during general elections and that cognitive dissonance is already low for people who simply vote for the first name, it seems still useful to control for these potential effects. For this reason, the candidates are presented in two different orders for each of the three groups. The order is chosen in such a way that the two middle candidates both are the first candidate once. Combining this with the control question to place the candidates in a political field it allows a check whether subjects simply vote for the first name.

The political placement of the four different candidates is based on five statements of each candidate about what they would implement if elected. Each candidate makes three economic statements and two statements about welfare and quality of life (Budge, 2013). The score of the statements of the candidates is based on the RILE-scale. This Right-Left scale can be used to scale ideological positions of candidates (Gabel & Huber, 2000). This scale is also used by the Manifesto Project, which determines political positions of parties to code over 1000 manifestos in more than 50 countries. The RILE-scale is used to construct the four candidates in such a way that one left-wing, one middle-left, one middle-right and one right-wing candidate can be identified. DeVries et al. (2013) provided evidence that most people are capable of placing candidates using such a unidimensional policy scale. The electorate is also capable of evaluating their political position and attitude in a particular policy field (Lo et al., 2014). Furthermore, the RILE-scale is constructed in such a way that the relative differences between the candidates equal each other. Appendix A gives a more detailed overview of the candidate descriptions.

3.3 Rating System

After reading the instructions and the statements of the candidates, the subjects have to rate the four different candidates. This experiment makes use of a thermometer rating scale which ranges from 0 to 100. The higher a subject rate a candidate, the more favourable the subject feels toward a particular candidate. Ratings below 50 degrees are considered as not favourable, and ratings above 50 degrees are considered as favourable. This thermometer rating scale is the most commonly used measure to evaluate candidates in a political field (Bar-Anan & Nosek, 2014).

After the first choice task, the subjects have to rate the candidates again. For the both treatment groups this choice task exists of the voting task, and for the control group, this choice task exists of the filler choice task. All the subjects were instructed that the second rating task was not a memory task to control for possible memory effects (Chartrand & Bargh, 2002). The difference in evaluative distance between the two rating tasks is used to measure the political polarisation. The spread of the evaluative distance between the chosen candidate(s) and the unchosen candidates of the first rating task need

to be compared with the same spread after the second rating task. If a subject rate the chosen candidate higher than the first time and/or rate the unchosen candidates lower than the first time, then this spread of evaluative distance has increased and can be considered as an increase in polarisation. This measure is often used to investigate choice induced attitude change in a political context (Duclos et al., 2004; Mullainathan & Washington, 2009; Elinder, 2012; McGregor, 2013).

$$\Delta\, polarisation = \Delta\, evaluative\ distance\ chosen\ candidate(s) - \Delta\, evaluative\ distance\ unchosen\ candidates$$

3.4 The Choice Tasks

All the subjects perform two different choice tasks. The control group starts with the filler choice task after the first rating task and finish with the vote-for-1 task after the second rating task. The first treatment group starts with the vote-for-1 task after the first rating test and end with the filler choice task after the second rating task. The second treatment group starts with the vote-for-2 task after the first rating task and end with the filler choice task after the second rating task. The filler choice task is constructed to equal the basic treatment flow of the three different groups. Every subject rate the candidates before doing a choice task and rate the candidates again before doing a second choice task.

The differences between the treatments are necessary to test both hypotheses. The first hypothesis tests whether the act of voting increases polarisation. For this reason, the only difference between the control group and the first treatment group is the moment in time they have to rate the candidates again. The control group rates the candidates again before the vote-for-1 task and the first treatment group after the vote-for-1 task. The second hypothesis tests whether a vote-for-2 system increases polarisation less than a vote-for-1 system. For this reason, the only difference between the first treatment group and the second treatment group is the different voting system.

Before the hypothetical election takes place, the subjects are noticed that their favourite candidate decided to withdraw from the election. This withdraw will force subjects to make a compromising choice, which creates cognitive dissonance. Literature shows that subjects are assumed to vote for their most preferred candidate. Weakening the link between the voting choice and the preference of a candidate is needed to use dissonance reduction as an instrument to influence candidate evaluations (Bølstad et al., 2013). There is no particular reason given of this withdrawing, to control for potential influences of these reasons (Rodriguez-Álvarez, 2006).

To equal all the treatment flows, a filler choice task is needed. For the control group, the filler choice task takes place before the voting task, and for both treatment groups, the filler choice task takes place after the voting task. This filler task is often used to control for memory effects (Crowder, 1967; Chartrand & Bargh, 2002). The filler task in this experiment needs to satisfy two conditions. First, it has to be a choice task where subjects can make a choice out of free will. Second, it has to cost subjects of the control group enough effort to control for memory effects between the two same rating tasks. For this reason, the filler task exists of choosing one option which appears most likely of three different instances. This task is used by Cohen & Chesnick (1972) to show that people tend to overestimate the probability of conjunctive events and underestimate the probability of disjunctive events.

3.5 Control variables

After finishing the two same rating tasks and two different choice tasks, all the subjects finish the experiment by answering eight questions which can be used as control variables. The first two questions serve as a purpose to place the different candidates and themselves in a political field from a left to a right wing. The last six questions are demographic questions to gather some extra control variables.

3.5.1 Political placement

First, subjects have to place the four candidates in a political field between the left wing and right wing. Subjects could place each candidate using an 11-point Likert scale based on Bakker et al. (2012) who reported the Chapel Hill Expert Surveys (2010), where 0 indicates a left-wing candidate, 5 indicates a middle candidate, and 10 indicates a right-wing candidate. Next, subjects have to place their own political beliefs using the same 11-point Likert scale. Subjects were asked: 'In political matters, people talk of "the left" and "the right." How would you place your views on this scale, generally speaking? [1 'Left' … 10 'Right']' (Lehmann, P., & Schultze, H., 2003).

The political placements serve several purposes. First of all, it can be used as a control variable. It can be used to check whether the precept saliency holds and it allows to control for the right placing of the candidates in a political field. As mentioned in Section 3.1.1 there is no monetary incentive to make 'good' choices. For this reason, the political placements can be used to check whether saliency is satisfied. If a subject rates the candidate who is the closest to their own political beliefs as most favourable, this can be considered as consistent. Furthermore, it allows checking whether subjects put enough effort in reading the instructions carefully. If a subject places the candidates close to the aimed political position, it confirms that a subject has read the instructions carefully and tried to answer

truthfully. Besides, if many subjects place one or more candidates at a different political position, this could influence the results. Lastly, it can be used to check whether subjects place themselves closer to the chosen candidate in a political field after the created dissonance due to withdrawing of their most favourite candidate.

3.5.2 Demographic questions

The demographic questions serve to gather additional control variables which could be of interest. Literature provides little evidence of control variables having an impact on political polarisation due to cognitive dissonance reduction. Mullainathan & Washington (2009) provided suggestive evidence that the level of interest of a person could affect dissonance reduction. For this reason, the dummy variable 'voted last election' is used as a proxy to test whether or not a subject is interested in politics. Bølstad et al. (2013) provided suggestive evidence that party membership also affects dissonance reduction. For this reason, also the dummy variable party membership is added as a control variable. Furthermore, the nationality of a subject is of importance to check whether a common voting system influences the evaluative distance. Lastly, the variables gender, age and highest achieved education are added. None of these variables are related to dissonance reduction based on literature. However, it could be of interest to see whether one of these variables influence the evaluative distance of the subjects.

3.6 Data

This section presents the data that has been gathered from the experiment. First, the raw data need to be cleaned before the descriptive statistics can be constructed. Section 3.6.1 explains the data cleaning process, and Section 3.6.2 describes the descriptive statistics of the final sample.

3.6.1 Data cleaning

The raw dataset is downloaded from the database in Qualtrics. The raw data exists of 157 participants who finished the survey, divided into three different groups. The raw data of the control group exists of 38 observations, the raw data of the first treatment group exists of 55 observations and the raw data of the second treatment group exists of 64 observations. This difference is made because of the expected higher variance in both treatment groups compared to the control group. For this reason, more observations are needed to equal the power of the different tests in Section 4. Before the gathered data is ready to analyse, it has to be cleaned. This means that several data points have to be excluded from the dataset for a particular reason.

After the first rating task, the subjects had to order the candidates from most preferred candidate to least preferred candidate. This ordering was necessary in Qualtrics to ensure that the most preferred candidate withdraws from the election which took place after the rating task. Subjects who ordered the candidates wrong are deleted from the sample because it is not sure whether the dissonance is created. For these subjects, their most favourite candidate did not withdraw from the election. For this reason, two subjects of the control group, seven subjects of the first treatment group and nine subjects of the second treatment group are deleted from the sample.

3.6.2 Descriptive statistics

The clean data set exists of 139 observations divided into three different groups. The control group exists of 36 observations, the first treatment group exists of 48 observations, and the second treatment group exists of 55 observations.

The total sample exists of 87(62.6%) males. The average age of the sample is 23.7 years old. Almost all subjects (94.2%) voted in the last election, and 10.1% of the subjects are a member of a political party. All the subjects have at least a high school degree, 87.8% of the subjects also have a bachelor degree and 34.5% of the subjects also have a master degree. The sample exists of three subjects who are allowed to vote for more than one country. The majority of the sample (88.5%) is allowed to vote in the Netherlands; only three subjects are allowed to vote in France; three subject in Germany; one subject in the UK; one subject in the US; and twelve subjects are allowed to vote in another country.

The most-left candidate Turquoise got an average rating of 31.87 points during the first rating task. The middle-left candidate Beige got an average rating of 41.14 points during the first rating task. The middle-right candidate Violet got an average rating of 70.54 points during the first rating task. The most-right candidate Peach got an average rating of 58.38 points during the first rating task. After the first choice task, all the subjects rated the candidates again. The average rating of Turquoise increased to 34.17 points, the average rating of Beige increased to 41.30 points, the average rating of Violet decreased to 69.70 points, and the average rating of Peach increased to 60.77 points.

During the voting task, the subjects in the control group and the first treatment group voted for one candidate. Candidate peach got 46 of the total 74 votes. Candidate Violet and candidate Beige both got 14 votes and candidate Turquoise got only 10 votes during this vote-for-1 task. The subjects in the second treatment group voted for two candidates. Candidate peach got 38 of the total 110 votes.

Candidate Beige got 35 votes, candidate Turquoise 20 votes and candidate Violet got only 17 votes during this vote-for-2 task.

The subjects in all groups were asked to place the candidates in a political field using an eleven point Likert scale where 0 indicates most left, and 10 indicates most right. On average, it seems that participants perceived the candidate positions as intended. The average indication of the subjects of candidate Turquoise is 2.47. The average indication of the subjects of candidate Beige is 3.18. The average indication of the subjects of candidate Violet is 6.96. The average indication of the subjects of candidate Peach is 7.82. The subjects place themselves on average at 6.19 using the same eleven point Likert scale. This own placement result seems consistent with the highest rating scores of candidate Violet because the placement result of candidate Violet (6.94) is the closest to the own average placement (6.19). Section 4.4 gives a more detailed explanation of these placement results. Table 2 provides an overview of the main descriptive statistics.

	Control group		Treatment group 1		Treatment group 2	
Subjects	36		48		55	
	Men	Vote last election	Member political party	High school degree	Bachelor degree	Master degree
Yes	87	131	14	139	122	48
No	51	8	125	0	17	91
	NL	France	Germany	UK	US	Other
Nationality[11]	123	3	3	1	1	12
	Rate 1	Rate 2	Single vote	Double vote	Placement	
Turquoise	31.87	34.17	10	20	2.47	
Beige	41.14	41.30	14	35	3.18	
Violet	70.54	69.70	14	17	6.96	
Peach	58.38	60.77	46	38	7.82	
Own political placement of subjects					6.19	

Table 2: Descriptive statistics of the total sample.

[11] Two subjects are allowed to vote for two countries and one subject is allowed to vote for three countries. For this reason the total amount of nationalities is 143 instead of 139.

4. Analyses and Results

This section gives an overview of the analyses and results. In Section 4.1 the two different measures of the spread of the evaluative distance, which is the dependent variable in this research, are described, and the descriptive statistics are given. In Section 4.2 several nonparametric tests and parametric tests are conducted to investigate the differences between the three groups. In Section 4.3 multiple regressions are carried out to interpret the results of the analyses. In section 4.4 some extra robustness checks are performed to investigate whether the intrinsic motivation of the subjects results in consistent answers. In section 4.5 both hypotheses are answered based on the results from this section.

4.1 Spread of the evaluative distance

In order to test the two hypotheses, the variable spread is created. Spread refers to the spread of the evaluative distance of the subjects. The variable spread is calculated by first taking the difference between the rating scores of the chosen candidate(s) and the unchosen candidates before the first choice task. For the control group and first treatment group, this score is calculated by taking the average of the three unchosen candidates and subtract it from the score of the chosen candidate. For the second treatment group, the mean of the two unchosen candidates is first calculated and then subtracted from the mean of the two chosen candidates. Then, the same difference is calculated after the first choice task (a filler task for the control group and a voting task for both treatment groups). By subtracting the first difference from the second difference, the variable spread is created. In this case, a positive spread means that a subject has increased the rating of the chosen candidate(s) and/or decreased the rating of the unchosen candidates and thus increased voting-induced polarisation.

Also, to secure robustness an additional way to calculate spread is conducted. Here, only the difference in rating scores of the third preferred candidate is computed. The assumption here is that the control group and first treatment group vote for their second most preferred candidate and the second treatment group vote for their second and third most preferred candidates. In fact, 127 subjects chose their second (and third) most favourite candidate as expected. For the other 12 subjects who did not choose their second (and third) most favourite, the rating scores are used as a so called ‘imputed choice’. This means that their ratings are used as if they voted for their second (and third) most favourite candidate. The difference in rating scores of the third preferred candidate (and the ‘imputed third preferred candidate) can be used to secure robustness concerning the choice attitude change of the subjects and could be useful to conclude on the second hypothesis. A positive spread here means

that the polarisation has increased and is only expected for the second treatment group because this is the only group who voted for the third preferred candidate. The difference with the main spread measure is, therefore, that the rating scores of the most preferred and least preferred candidate are ignored here. This measure of spread is referred to as 'alternative measure of spread' during this research.

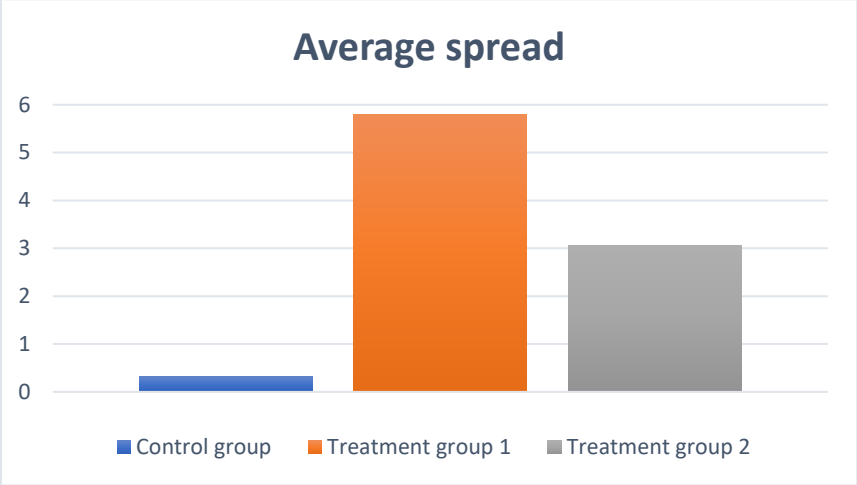

Figure 2: Average spread of each group

Figure 2 shows the average spread of each group. The average spread of the control group is 0.315 points. This result means that, on average, a subject in the control group increases the difference in rating score between the chosen candidate and the average of the unchosen candidates with 0.315 points after the first rating task. Here, the choice was made after both rating tasks. Since no choice is made between ratings, one would not expect any spread. The average spread of the first treatment group is 5.799 points. This result means that, on average, a subject in the first treatment group increases the difference in rating score between the chosen candidate and the mean of the unchosen candidates with 5.799 points after the choice for one candidate. Here, the choice for one candidate was made between the two rating tasks. Therefore, one would expect a larger spread compared to the other two groups. The average spread of the second treatment group is 3.064 points. This result means that, on average, a subject in the second treatment group increases the difference in rating score between the mean of the two chosen candidates and the average of the two unchosen candidates with 3.064 points after voting for two candidates. Here, the choice for two candidates was made between the two rating task. Therefore, one would expect a larger spread compared to the control group and a smaller spread than the first treatment group

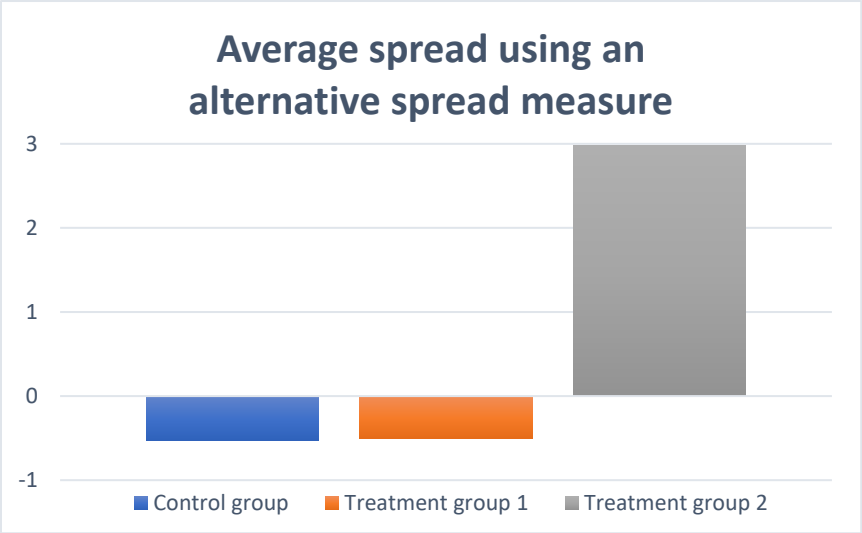

Figure 3: Average spread of each group using the alternative measure of spread

Figure 3 shows the spread of each group using the alternative measure of spread. Here, only the difference in rating scores of the third preferred candidate is calculated. The average of the spread of the control group is -0.528 points. This result means that, on average, a subject in the control group decreases the thermometer rating score of their third preferred candidate with 0.528 points after choosing their second preferred candidate. The average of the spread of the first treatment group is -0.500 points. This result means that, on average, a subject in the first treatment group decreases the thermometer rating score of their third preferred candidate with 0.500 points after choosing their second preferred candidate. The average of the spread of the second treatment group is 2.982 points. This result means that, on average, a subject in the second treatment group increases the thermometer rating score of their third preferred candidate with 2.982 points after choosing their second and third preferred candidate.

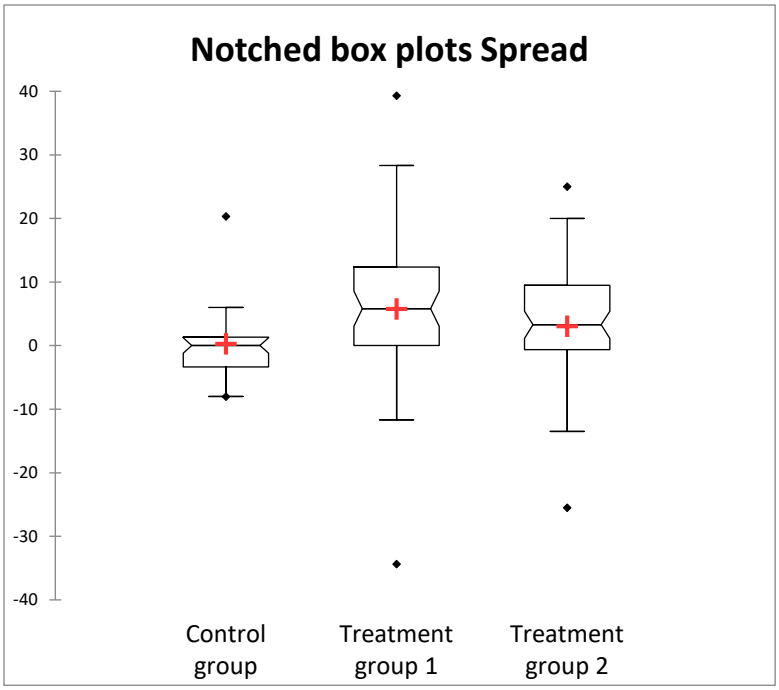

Figure 4: Notched box plot spread of the evaluative distance

Figure 4 shows a notched box plot of the spread of the evaluative distance for all the three groups. The box shows the interquartile range (IQR), and the whiskers subtract and add 1.5 times the IQR from the first and third percentile. The line shows the median and the red plus illustrates the mean of the data. The notches represent 95% confidence intervals for the estimated medians. If two notches from different boxes do not overlap, there is strong evidence their median differ (using a 95% confidence interval). These notches seem to provide evidence that the medians of the control group and first treatment group differ.

Furthermore, the box plot of the control group exists of a mean of 0.315, a standard deviation of 5.884, a standard error of 0.981 and a 95% confidence interval between -1.607 and 2.237. This result of almost no spread is in line with the expectations because only a filler choice is made between the two rating tasks. The boxplot of the first treatment group exists of a mean of 5.799, a standard deviation of 13.082, a standard error of 1.888 and a 95% confidence interval between 3.701 and 9.499. This spread that is larger than the spread of the control group is in line with the expectations because the literature provides evidence for this choice induced attitude change after voting for a candidate. The boxplot of the second treatment group exists of a mean of 3.064, a standard deviation of 9.665, a standard error of 1.303 and a 95% confidence interval between 2.554 and 5.618. This spread that is larger than the spread of the control group but lower than the spread of the first treatment group is in line with the expectations because choice induced attitude change is expected here, but less than in the first treatment group. This can be explained by the fact that people may feel less urge to vote

strategically or compromisingly when they are allowed to vote for more than one candidate compared to a situation in which subjects vote for one candidate. An overview of the descriptive statistics of the spread variable is presented in table 3.

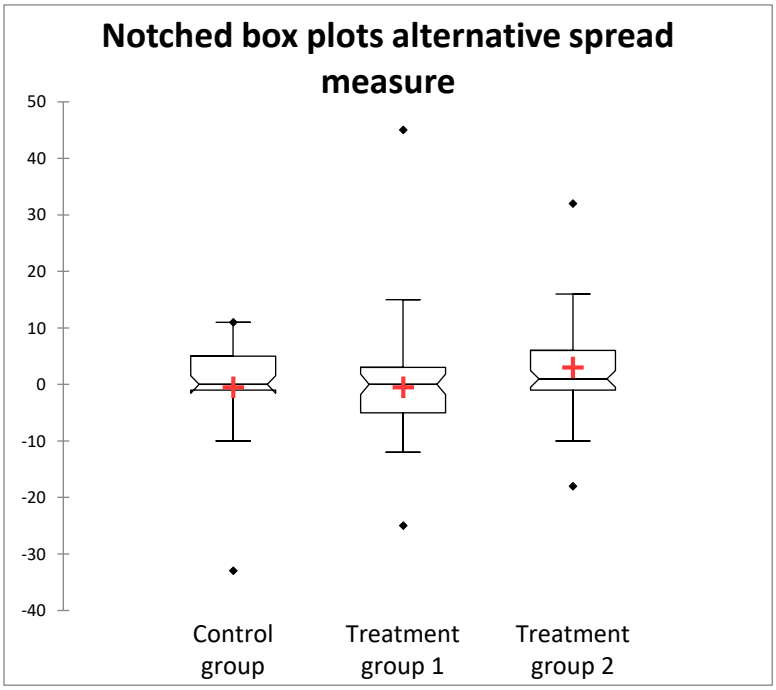

Figure 5: Notched box plot alternative measure of spread

Figure 5 shows a notched box plot of the spread of the evaluative distance for all the three groups using the alternative measure of spread. Here, the notches seem to overlap in all three box plots which means that the medians do not differ using a 95% confidence interval. Furthermore, the result of almost no spread in the control group is in line with the expectations because only a filler choice is made between the two rating tasks. The result of almost the same mean in the first treatment group as in the control group is a little surprising. Contrary to the control group, the first treatment group shows that the third preferred candidate is a rejected candidate during the second rating task. Therefore, according to the literature about choice induced attitude change after rejecting an alternative, one should expect a larger decrease compared to the control group. However, the box plot as a whole indeed results in a little lower spread in the first treatment compared to the control group. This result of a positive spread in the second treatment group is in line with the expectations because only in the second treatment group the third preferred candidate is a chosen alternative. The statistics of this measure of spread are also given in table 3.

	Mean	Std. dev.	Std. error	Median	Kurtosis	Skew-ness	Min	Max	Obs.
Control A	0.315	5.884	0.981	0.000	3.311	1.462	-8.000	20.333	36
TG1 A	5.799	13.082	1.888	5.500	2.677	-0.602	-34.333	39.333	48
TG2 A	3.064	9.665	1.303	3.500	0.925	-0.529	-25.500	25.000	55
Control B	-0.528	8.567	1.428	0.000	5.122	-1.769	-33.000	11.000	36
TG1 A	-0.500	10.542	1.506	0.000	6.702	1.305	-25.000	45.000	48
TG2 B	2.982	10.356	1.384	1.000	1.207	0.841	-18.000	32.000	55

Table 3: Descriptive statistics for spread variables

Table 3 shows an overview of all the statistics of the calculated spread variable (A). This dependent variable is needed to answer the both hypotheses and the research question. Also, the descriptive statistics of the alternative spread variable (B) for extra robustness are given in table 3.

4.2 Non-parametric and parametric tests

To provide statistical evidence for the indicated results of the previous section, several statistical tests are conducted. This section presents nonparametric tests to conclude on the samples, and some parametric tests to ensure robustness. To conclude on parametric tests, four assumptions are needed. The observations need to be independent, they must be drawn from a normally distributed population, two groups must have the same variance, and the variables must be measured on an interval scale (Glass, Peckham, & Sanders, 1972). The observations are independent and on an interval scale. However, the observations are not drawn from a normally distributed population[12], and the samples do not have the same variance. The two-sample t-test can be done using an unequal variance of the data to control for this problem, but the missing normally distribution cannot be tackled properly. For this reason, these results are only used to ensure robustness, but not to conclude on the hypotheses. For non-parametric tests, only the assumption of independent observations needs to hold in this research (Glass, Peckham, & Sanders, 1972). For this reason, the conclusions are drawn based on the nonparametric tests and the t-tests are added for extra robustness of the results.

4.2.1 Individual sample tests

First, each sample is individually compared with a created sample of only zero's. The non-parametric Wilcoxon signed-rank test is used to test whether the three different groups individually differ from zero. For the control group, the result is W = 176 with a p-value of 0.764. This means that the null

[12] Several normality tests are done to check whether the spread variable follows a normal distribution. The results of the Jarque-Bera tests provide evidence that the null hypothesis that the sample shows a normal distribution can be rejected for all the three samples at a 1% significance level.

hypothesis that the median spread in the control group is equal to zero cannot be rejected at a 10% significance level[13]. This result is in line with the expectations of no attitude change and the results of Section 4.1. For the first treatment group, the result is W = 828 with a p-value of 0.000. This means that the null hypothesis that the first treatment group is equal to zero can be rejected at a 1% significance level. This result is in line with the expectations of a choice induced attitude change and the results of Section 4.1. For the second treatment group, the result is W = 979.5 with a p-value of 0.008. This means that the null hypothesis that the second treatment group is equal to zero can be rejected at a 1% significance level. Based on the literature about choice induced attitude change this outcome was hard to predict. Nevertheless, this result is in line with the results of Section 4.1.

Next, in order to secure robustness, three one-sample t-tests are done to test whether the three different groups individually differ from zero. For the control group the result is: t(35) = 0.321 with a p-value of 0.750. This means that the null hypothesis that the mean of the control group equals zero cannot be rejected at a 10% significance level. This result is in line with the non-parametric test. For the first treatment group, the result is t(47) = 3.071 with a p-value of 0.004. This means that the null hypothesis that the mean of the first treatment group equals zero can be rejected at a 1% significance level. This result is in line with the nonparametric test For the second treatment group the result is: t(54) = 2.351 with a p-value of 0.022. This means that the null hypothesis that the mean of the first treatment group equals zero can be rejected at a 5% significance level. This result is in line with the non-parametric test.

4.2.2 Multiple samples tests

The non-parametric Kruskal-Wallis test is conducted to test whether the three different samples come from the same population. The result of the Kruskal-Wallis is $\chi^2(2) = 11.400$ with a p-value of 0.003. This means that the null hypothesis that the samples come from the same population can be rejected at a 1% significance level. To test whether two of the three samples come from the same population, three Mann-Whitney U tests are conducted. The result of the Mann-Whitney U test between the control group and the first treatment group is: $U = 500.500$ and Z = -3.286 with a p-value of 0.001. This means that the null hypothesis that the difference in the distribution of spreads between the samples is equal to zero can be rejected at a 1% significance level. The result of the Mann-Whitney U test between the control group and the second treatment group is: $U = 705.500$ and Z = -2.309 with a p-

[13] This thesis makes use of three different levels of significance. All conclusions are based at a 5% level. This means that every p-value which is higher than 5% is rejected. To indicate the level of significance a 1% and a 10% level is also mentioned.

value of 0.021. This means that the null hypothesis that the difference between the samples is equal to zero can be rejected at a 5% significance level. The result of the Mann-Whitney U test between the first treatment group and the second treatment group is *U* = 1522.500 and Z = 1.339 with a p-value of 0.182. This means that the null hypothesis that the difference between the samples is equal to zero cannot be rejected at a 10% significance level[14].

Next, in order to secure robustness also three parametric two sample t-tests with unequal variances are conducted. The result of the two sample t-test between the control group and the first treatment group is t(69) = -2.577 with a p-value of 0.012. This means that the null hypothesis that the difference between the samples is equal to zero can be rejected at a 5% significance level. This result is in line with the earlier finding from the nonparametric test. The result of the two sample t-test between the control group and the second treatment group is t(89) = -1.685 with a p-value of 0.095. This means that the null hypothesis that the difference between the samples is equal to zero cannot be rejected at a 5% significance level. This result conflicts with the earlier finding from the nonparametric test. The result of the two sample t-test between the first treatment group and the second treatment group is t(86) = 1.192 with a p-value of 0.237. This means that the null hypothesis that the difference between the samples is equal to zero cannot be rejected at a 10% significance level. This result is in line with the earlier finding from the nonparametric test.

4.2.3 Alternative spread measure

The same tests that have been used for the main measure of spread are conducted using the alternative measure of spread. Conducting the same Wilcoxon signed-rank tests as before, it cannot be rejected for both for the control group and first treatment group that the mean differs from zero at a 10% significance level (p-values: 0.753 and 0.450). For the second treatment group, this can be rejected at a 10% level, but also not at a 5% level (p-value: 0.059). After conducting three parametric one sample t-tests, it cannot be rejected for both for the control group and first treatment group that the mean differs from zero at a 10% significance level (p-values: 0.746 and 0.714). For the second treatment group, this can be rejected at a 5% significance level (p-value: 0.039). The majority of these results are in line with the literature and the results in Section 4.1. Nevertheless, interpreting the Wilcoxon signed-rank test, the second treatment group does not significantly differ from zero. This result conflicts with the expectations and results of Section 4.1.

[14] Adjusting p-values to correct for family-wise error rates with the Holm-Bonferroni adjustment procedure does affect the conclusions drawn.

Conducting a Kruskal-Wallis test using this alternative measure of spread, one cannot reject that the samples come from the same population at a 10% significance level (p-value: 0.169). To review the differences between the samples the same Mann-Whitney U tests were conducted as before. Here, all the null hypotheses that the difference between the samples equals zero cannot be rejected at a 5% significance level (p-values are CG-TG1: 0.481, CG-TG2: 0.297 and TG1-TG2: 0.066). The two-sample t-test results confirm these results. This result is somewhat surprising. Based on the literature one would expect a significant difference between the both treatment groups. This p-value is significant at a 10% level, but not at the 5% level. Appendix B gives an overview of the results of all the test using the alternative measure of spread.

4.3 Regressions

To check the significance of the results found in Section 4.1 and 4.2, multiple regression models are conducted. For the main conclusions, section 4.3.1 shows the detailed results of the main spread measure. To ensure robustness section 4.3.2 shows the basic results of the alternative spread measure.

4.3.1 Spread

For the regression models in this section, the dependent variable is the spread variable which is calculated by taking the difference in rating scores between the chosen and unchosen candidates. In the first regression model, a regression is conducted using both treatment groups as the only independent variables. Then, in the second, third, fourth and fifth model some control variables are added individually. In the sixth model, all the control variables are added together in the same regression. After running the first regression, several tests are done to check whether the OLS regression satisfies the Gauss-Markov assumptions needed to be the Best Linear Unbiased Estimator. The Breusch-Pagan/Cook-Weisberg test shows that this regression suffers from heteroscedasticity (p value = 0.000). For this reason, only robust standard errors are used to tackle this problem. After this correction, all Gauss Markov conditions for OLS to be the best linear unbiased estimator are satisfied. The results of the six different models are presented in table 4.

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Treatment group 1	5.484*** (2.127)	5.339** (2.152)	5.151** (2.184)	5.121** (2.232)	5.540*** (2.133)	4.794** (2.335)
Treatment group 2	2.275* (1.631)	2.790* (1.647)	2.719* (1.619)	2.584 (1.647)	2.746* (1.654)	2.593 (1.677)
Man	No	-2.599 (1.786)	No	No	No	-1.975 (1.803)
Dutch voter	No	No	-3.988 (2.821)	No	No	-3.583 (3.010)
Member Political party	No	No	No	2.611 (2.587)	No	2.474 (2.621)
Master degree	No	No	No	No	1.159 (1.802)	1.269 (1.784)
Constant	0.315 (0.978)	1.975 (1.493)	3.971 (2.954)	0.242 (0.989)	-0.104 (1.220)	4.334 (3.225)
Observations	139	139	139	139	139	139
F-score	3.84** (10.254)	2.86** (10.213)	4.16*** (10.212)	3.30** (10.262)	2.65* (10.277)	2.24** (10.228)
R-squared	0.042	0.057	0.057	0.047	0.045	0.075

Table 4: Regression models spread

***p<0.01, **p<0.05, *p<0.10
(Robust standard errors)

The first treatment group shows significant results in every model at a 5% significance level and in model one and five even at a 1% significance level. The first model illustrates that, on average, a subject in the first treatment group shows an increase in the spread between chosen and unchosen candidates, comparing the first and second rating task, of 5.584 points compared to a subject in the control group, ceteris paribus. This increase is significant at a 1% significance level. The second treatment group does not show significant coefficients at a 5% significance level. However, in model one, two, three and five, it shows significant results at a 10% significance level. Interpreting the first model once more, this means that, on average, a subject in the second treatment group shows an increase in the spread between chosen and unchosen candidates between the first and second rating task of 2.275 points compared to a subject in the control group, ceteris paribus. However, this result is not significant at a 5% significance level.

In model two, three, four and five some control variables are added individually. Not all the control variables resulting from the experiment have been used. Some control variables are dropped due to an extremely low variance, which made them irrelevant. The control variable nationality is changed to a dummy Dutch/non-Dutch and the control variable highest achieved degree is changed to a dummy Master/non-master degree. All the control variables show insignificant coefficients both on the

individual level as jointly. The gender dummy variable 'man' is with a p-value of 0.148 the least insignificant control variable.

Lastly, an additional regression is executed to test the difference between the first treatment group and the second treatment group. The coefficient of this regression is -2.735 with a robust standard error of 2.296. However, this result is not significant at a 10% significance level and therefore not further interpreted in this section.

4.3.2 Alternative spread measure

In order to secure robustness of the results found in Section 4.3.1, several regression models are conducted using the alternative spread measure. For these regression models, the dependent variable is the alternative spread variable which is calculated by taking the difference in rating scores between the third preferred candidate. In the first regression model, a regression is conducted using only the second treatment group as independent variables. In the second model, both treatment groups are added as an independent variable. In the third model, the second treatment group and the same control variables as in Section 4.3.1 are added together as independent variables. Lastly, the both treatment groups and the control variables are added together as independent variables in model 4. After running the first regression, several tests are done to check whether the OLS regression satisfies the Gauss-Markov assumptions mentioned in Section 4.3.1. The Breusch-Pagan/Cook-Weisberg test shows that this regression does not suffer from heteroscedasticity (p value = 0.583). In addition to this, the other Gauss Markov conditions for OLS to be the best linear unbiased estimator are satisfied in this regression. The results of the four different models are presented in table 5.

Independent variables	Model 1	Model 2	Model 3	Model 4
Treatment group 1	No	0.028 (2.221)	No	-0.365 (2.277)
Treatment group 2	3.494** (1.741)	3.510 (2.160)	3.695** (1.750)	3.490 (2.172)
Man	No	No	-0.178 (1.823)	-0.179 (1.830)
Dutch voter	No	No	-2.834 (2.764)	-2.881 (2.789)
Member Political party	No	No	0.021 (2.857)	0.108 (2.918)
Master degree	No	No	-2.611 (1.802)	-2.623 (1.811)
Constant	-0.512 (1.095)	-0.528 (1.679)	2.927 (2.674)	3.172 (3.087)
Observations	139	139	139	139
F-score	4.03** (10.037)	2.00 (10.074)	1.51 (10.055)	1.25 (10.092)
R-squared	0.022	0.014	0.018	0.012

Table 5: Regression models of alternative spread measure

***p<0.01, **p<0.05, *p<0.10
(Standard errors)

Table 5 shows significant coefficients for the second treatment group in model one and three at a 5% significance level. The first model illustrates that, on average, a subject in the second treatment group shows an increase in the spread between the third preferred candidate, comparing the first and second rating task, of 3.494 points compared to a subject not in the second treatment group, ceteris paribus. This increase is significant at a 5% significance level. Comparing both treatment groups individually to the control group does not show significant coefficients at a 10% significance level. Again, all the control variables show insignificant coefficients. The master dummy variable is with a p-value of 0.150 the least insignificant control variable.

4.4 Robustness checks

Before concluding on the hypotheses, it is important to investigate the robustness of the obtained results of the previous sections. As described in Section 3.1.1, this experiment does not satisfy the precepts of Smith (1982) which are needed for a controlled economic experiment because of a leak of incentives. This means that the subjects mostly answer the questions based on their intrinsic motivation. Although Vossler & Kerkvliet (2003) and Delavande & Manski (2015) already provided evidence that political choices in surveys or experiments match actual voting outcomes because of the

importance of elections, a closer look at the answers to some of the control questions is needed to test whether this is also the case this experiment.

4.4.1 Political placement

First, subjects were asked to place the candidates in a political field on a 0-10 scale, where 0 indicates a far-left candidate, and 10 indicates a far-right candidate. The political placement of the four different candidates is created by five statements of each candidate about what they would implement if elected. The score of the statements of the candidates is based on the RILE-scale. This Right-Left scale can be used to scale ideological positions of candidates (Gabel & Huber, 2000). The RILE-scale is used to construct the four candidates in such a way that one left-wing, one middle-left, one middle-right and one right-wing candidate can be identified.

About 93% of the subjects were able to indicate the difference between the two right candidates and two left candidates. This result means that almost 93% of the subjects placed the two left candidates left from the two right candidates. However, only 54% of the subjects were able to place all the candidates in the right order. Almost 46% of the subjects placed one or two middle candidate(s) more to the wing than the wing candidate(s). This potential problem is further investigated by taking the average outcomes of the placements. The most-left candidate is placed on an average of 2.47, while an average of 2 was expected using the intended RILE scale. The middle-left candidate is placed on an average of 3.18, while an average of 4 was expected using the intended RILE scale. The middle-right candidate is placed on an average of 6.96, while an average of 6 was expected using the intended RILE scale. The most-right candidate is placed on an average of 7.82, while an average of 8 was expected using the indented RILE scale. The relative distances between the candidates were intended for two points between each candidate on the RILE scale. The real relative distances between the candidates are from left to right respectively; 0.71, 3.78 and 0.86. Next, the distance between the two wing candidates and the minimum and maximum of the RILE score was also intended for two points. The real differences are 2.47 from the minimum and 2.18 from the maximum postion. These results show that subjects, on average, were able to place the candidates in the right order from left to right. However, the two middle-candidates seem to be placed too far to the wings. It is important to investigate whether this influence the results.

After the subjects had placed the candidates, they had to place themselves on the same RILE scale. The average score for the subjects is 6.19. By creating a proxy for created dissonance, the influence of the difference in RILE scores can be investigated. The proxy for dissonance is calculated by taking the

difference between the own political placement of subjects and the political placement of the chosen candidate. The result of the proxy for dissonance using the indented RILE score is 1.597 (6.19 – average intended RILE score from the chosen candidates), and the result of the proxy for dissonance using the real RILE score is 1.669 (6.19 – average real RILE score from the chosen candidates). These results show that, because of the difference in real placement scores and the intended benchmark, the created dissonance is on average 0.072 points lower than the intended RILE scores should predict. Using a t-test one cannot reject that the difference between these samples equals zero at a 10% significance level (p-value: 0.792). To conclude, the placement of the candidates is something to keep in mind while concluding, but it does not influence the created dissonance significantly when using this proxy for cognitive dissonance.

4.4.2 Voting behaviour

As an additional robustness check, this subsection investigates the individual voting behaviour. Logically, one would expect that subjects in the control group and first treatment group vote for their second preferred candidate and subjects in the second treatment group for their second and third preferred candidate. The results show that in the control group and first treatment group 92.86% of the subjects voted for their second preferred candidate and in the second treatment group 89.09% voted for their second and third preferred candidate. These results can both be considered as an indication that the majority of the subjects have thought hard enough about their voting decision to vote consistent with their rating scores.

Another way to analyse the consistency of subjects is by checking whether they vote for the remaining candidate which they indicated as closest to their own political placement. For the control group and first treatment group, 89.29% of the subjects voted for the remaining candidate which they indicated as closest to their own political beliefs. For the second treatment group, 76.36% of the subjects voted for the two remaining candidates which they indicated as closest to their own political beliefs. These huge figures seem to provide evidence that thermometer ratings in large reflect underlying political preferences. However, important here is to mention that subjects indicated their own political beliefs after they have rated, voted and placed the candidates. This outcome indicates that subjects are consistent in their answers, but not necessarily whether they voted for the 'right' candidate. This consistency leaves two possible explanations. First, subjects were honest and thus consistent. Second, subjects were inattentive in the beginning and then exerted a high level of mental effort to remember all their answers to achieve consistency. Here, the first explanation seems more realistic, and therefore the consistency appears to be the result of honest answers.

It is interesting to see the fact that 21.58% of the subjects placed their own political beliefs closer to the remaining candidate(s) for which they voted than for the candidate they had indicated as their most favourable candidate before the voting took place. Also, 22.30% of the subjects placed their own political beliefs equally close to the remaining candidate(s) for which they voted as for the candidate they had indicated as their most favourable candidate. This result could be an indication of dissonance reduction. However, it is impossible to provide statistical evidence using this experiment.

4.4.3 Other checks

This subsection investigates some other small robustness checks to control for the missing extrinsic motivation during this experiment. First, the average time a participant spent to complete this experiment was 9 minutes and 33 seconds with a standard deviation of 4 minutes and 38 seconds[15]. Before the experiment went online, five persons tried the experiment to indicate the average time. The average duration of these tests was 9 minutes and 4 seconds. Furthermore, every participant was informed during the introduction text that the experiment should take approximately 10 minutes of their time. These results, therefore, indicate that subjects spend enough time to read the instructions and answer the questions carefully.

Another obtained result of interest is the change in total thermometer rating score of a subject. One could expect that subjects rate a candidate on average around 50 points because they compare the candidates to each other. This should mean that the average rating of a candidate should not change significantly for the second rating task. The results show that on average a subject rate a candidate with 50.48 points the first rating task and with 51.48 the second rating task. The hypothesis that the outcomes of the two samples follow the same distribution can be rejected at a 5% significance level using a Wilcoxon test (p-value: 0.011). This means that, on average, a subject rate the candidates one point higher during the second rating task. For the main measure of the spread of this research, this is not a problem because the distance is measured relatively between the chosen and unchosen alternatives. However, for the alternative measure of spread, this means that the results of the increased spread during the second task have to adjust a little bit because on average a subject rate a candidate already 1 point higher. Nevertheless, when comparing the samples with each other, there is already controlled for this effect.

[15] A total of 8 subjects were not added to the calculations because they spend more than 30 minutes over the average duration time, which indicates that that they have paused the experiment and continued after some time.

4.5 Hypotheses

All analyses have been performed in order to conclude on the two hypotheses of this research. This section summarises the main findings to conclude on both hypotheses.

4.5.1 Hypothesis 1

The first hypothesis stated that '*A voting system in which people vote for one candidate increases the polarisation of the electorate'.* To conclude on this hypothesis, the first treatment group has to be compared with the control group. The only difference between these groups is that the first treatment group vote for a candidate before the second rating task while the control group vote after the second rating task.

First, the Wilcoxon signed-rank test and the one sample t-test show that the spread of the evaluative distance in the control group does not differ significantly from zero. This result is important to ensure that the obtained differences between the two groups come from the choice induced attitude change and not from something else. The same Wilcoxon signed-rank test and t-test show that the spread of the evaluative distance in the first treatment group does differ significantly from zero. On average, the spread of the evaluative distance is 5.799 in the first treatment group. The t-score of the one sample t-test is 3.071, and this effect is significant at a 1% significance level for both the Wilcoxon signed-rank test and the one sample t-test. This result is a first indication that this hypothesis cannot be rejected.

The check whether the two groups significantly differ from each other a Mann-Whitney U test is performed, and an additional two sample t-test is used for extra robustness. The U score of this test is 500.500, with a Z-score of -3.286 and a p-value of 0.001. This means that the null hypothesis that the difference between the samples is equal to zero can be rejected at a 1% significance level. The result of the additional two sample t-test between the control group and the first treatment group is t(69) = -2.577 with a p-value of 0.012. This p-value confirms that the null hypothesis that the difference between the samples is equal to zero can be rejected at a 5% significance level. This result is a second indication that this hypothesis cannot be rejected.

Lastly, some regressions are performed to conclude on this hypothesis. The first OLS regression without all the insignificant control variables provides evidence that on average the spread of the evaluative distance in the first treatment group is 5.484 larger than the same spread in the control

group. This effect is significant at a 1% significance level. Concluding, the first hypothesis of this research cannot be rejected.

4.5.2 Hypothesis 2

The second hypothesis stated that '*A voting system in which people vote for two candidates increases the polarisation of the electorate relatively less than a voting system in which people vote for one candidate'.* To conclude on this hypothesis, the second treatment group has to be compared with the first treatment group. The only difference between these groups is that the second treatment group vote for two candidates and the first treatment group vote for one candidate. As reviewed in Section 4.5.1, the first treatment group shows a significant increase in the spread of the evaluative distance. To conclude on this hypothesis, the same measure of spread is used for the second treatment group, and one additional measure of spread is included. The alternative measure of spread uses the difference in rating scores of the third preferred candidate and compare this result between the groups.

First, Wilcoxon signed-rank test and the one sample t-test show that the spread of the evaluative distance in the second treatment group does differ significantly from zero. On average the spread of the evaluative distance is 3.064 in the second treatment group. The t-score of the one sample t-test is 2.351, and this effect is significant at a 5% significance level using a one sample t-test and at a 1% significance level using the Wilcoxon signed-rank test. This result indicates that the second treatment group shows a significant increase of spread which is lower compared to the first treatment group. However, this is not enough evidence to conclude on this hypothesis.

To check whether the two groups significantly differ from each other a Mann-Whitney U test is performed, and one additional two sample t-test is used for extra robustness. The U score of this test is 1522.500, with a Z-score of Z = 1.339 and a p-value of 0.182. This means that the null hypothesis that the difference between the samples is equal to zero cannot be rejected at a 10% significance level. The same Mann-Whitney U test between the control group second treatment group resulted in a U score of 705.500, with a Z-score of -2.309 and a p-value of 0.021. This means that the null hypothesis that the difference between the samples is equal to zero can be rejected at a 5% significance level. The result of the additional two sample t-test between the first treatment group and the second treatment group is t(86) = 1.192 with a p-value of 0.237. This result confirms that the null hypothesis that the difference between the samples is equal to zero cannot be rejected at a 10% significance level. The result of the additional two sample t-test between the control group and the second treatment group

is t(89) = -1.685 with a p-value of 0.095. This conflicts with the non-parametric outcome that the null hypothesis that the difference between the samples is equal to zero can be rejected at a 5% significance level. Again, more analyses are needed to conclude on this hypothesis.

An additional measure of spread can be used to ensure robustness for this hypothesis. Using a one-sample t-test for this measure of spread, one can conclude that the mean of first treatment group do not significantly differ from zero while the mean of the second treatment group does significantly differ from zero. However, using a Mann-Whitney U test the null hypotheses that the difference between the samples equals zero cannot be rejected at a 5% significance level. Also, the null hypotheses that the difference between the second treatment group and the control group equals zero cannot be rejected at a 5% significance level. This is in line with the result of the main measure of spread used for this hypothesis. Both results indicate that the difference in the spread of the evaluative distance between the first treatment group and the second treatment is not significant. However, more analyses are needed to conclude on this hypothesis.

Lastly, some regressions are performed to conclude on this hypothesis. The first OLS regression without all the insignificant control variables provides evidence that on average the spread of the evaluative distance in the second treatment group is 2.275 larger than the same spread in the control group. This effect is significant at a 10% significance level, but not at a 5 % significance level. This means that one cannot reject that voting for two candidates increases the polarisation of the electorate. To conclude whether voting for the two candidates increases the polarisation of the electorate less than voting for one candidate the same OLS regression is conducted to compare the both treatment groups with each other. The result of this regression shows a decrease in the spread of the second treatment group compared with the first treatment group of 2.735 points. However, this result is not significant at a 10% significance level. Concluding, the second hypothesis of this research needs to be rejected.

Summarising, a voting system in which people vote for one candidate increases the polarisation of the electorate. Next, it cannot be concluded at a 5% significance level that a voting system in which people vote for two candidates increases the polarisation of the electorate less than voting for one candidate.

5. Conclusion

This thesis investigated a potential causal relationship between different voting systems and polarisation. In order to investigate this relationship, an online survey is created in which 139 subjects participated in a hypothetical election. The subjects were randomly divided into three different groups. In the first treatment group subjects faced a common voting system where they had to vote for one candidate. In the second treatment group subjects were confronted with an alternative voting system where they had to vote for two candidates. To create cognitive dissonance the most preferred candidate of each subject withdraws just before the election took place. Subjects had to rate the candidates before and after the election to observe the change in polarisation. By taking the difference in evaluative distance between the chosen and unchosen candidates the proxy spread was calculated. This spread is compared between the two treatment groups and with the control group in which subjects rated the candidates two times before the voting task. The obtained results show that a voting system in which people vote for one candidate increases polarisation, but it cannot be concluded at a 5% significance level that a voting system in which people vote for two candidates increases polarisation less than a voting system in which people vote for one candidate.

5.1 Discussion

This thesis contributes to the existing literature in two fold. First, this research uses an experiment to investigate the relationship between voting-induced attitude change and polarisation. Existing literature only provides analyses with historical data. Second, this research contributes by investigating the influence of different voting systems on voting-induced attitude change. By forcing people to vote for two candidates instead of one, the voting-induced attitude change could be lower because people can vote less strategically and compromisingly which result in less cognitive dissonance. Despite these contributions, the results are not without flaws.

First of all, this experiment does not satisfy all the precepts of Smith (1982) which are needed to obtain control. Incentivizing subjects on a question level was simply not possible using this experimental design. The only monetary incentive, in this case, encouraged people to finish the experiment, but not to answer each question truthfully. For these reasons, subjects need to be intrinsically motivated to answer truthfully. Several tests are done to check the consistency of the answers. The results of these tests show that most subjects answer consistently. However, this does not necessarily mean that subjects also have answered honestly. The assumption here is that subjects who did not answer

honestly will not put effort into trying to 'lie' consistently during the entire experiment. Nevertheless, performing this experiment with monetary incentives could lead to more control and therefore more robust results. Furthermore, the experiment was not executed in a controlled lab environment, but by using an online link, which also results in a leak of control. Future research could tackle this problem by conducting an experiment in a controlled lab environment or a framed field environment[16].

Secondly, the results of this research show some indications that not enough subjects participated in the experiment to ensure the right power for the analyses. Because of the expected difference in variance between the both treatment groups compared to the control group, more subjects were needed in these treatments to ensure enough power. The randomizer in Qualtrics was set in such a way that participants had 40% chance to participate in treatment 1, 40% in treatment 2 and 20% in the control group. However, due to several constraints and the data exclusions, these deviations differed from the expectations. The results of the second treatment group had a p-value of 0.094. A mean of 3.064 and a high standard deviation of 9.665. Based on 55 participants, these numbers could indicate that more subjects may result in significant results. Future research, therefore, could make sure that more subjects participate in the second treatment group to investigate this speculation in more detail.

Thirdly, the descriptions of the candidates used in this experiment have left room for improvement for future research. Section 4.4 already showed that 93% of the subjects were able to indicate the difference between the two right candidates and two left candidates. However, only 54% of the subjects were able to place all the candidates in the correct order. Here, the two middle-candidates were placed more to the wings than the intended RILE score predicted. After using a proxy for created dissonance, these placement differences resulted in 0.072 points less created dissonance on an 11 point Likert scale, than the intended RILE scores predicted. However, this difference was not significant at a 10% level. Nevertheless, the difference in placement of the candidates is something to keep in mind. One cannot know for sure what and how much a different placement affects other answers of subjects. Furthermore, the proxy for created dissonance is not based on literature but based on assumptions. In order to improve these results, future research should find a way to place the two middle candidates clearly more to the middle than the both wing candidates. The easiest way to achieve this is including more than five statements.

[16] A framed field experiment can be described as an experiment with a nonstandard subject pool and with field context in either the commodity, task, or information set that the subjects can use (Harrison & List, 2004).

Fourthly, the time passage between the two rating tasks is relatively small. The both treatment groups have to vote between the two rating tasks, while the control group perform a filler choice task between the rating tasks. This leads to a setting in which subjects do not have much time between the two rating tasks. The voting task and filler task both required full attention and the subjects were wanted not to consider the task as a memory game. However, due to the small amount of time, subjects might have examined the task still as a memory game. Future research could tackle this problem by including more time and/or filler tasks between the two rating tasks. Furthermore, because of the two rating tasks in the same experiment, only a short time effect is obtained. In order to conclude on voting-induced attitude change in the long run, the experiment should be expanded with multiple periods over a couple of years.

Lastly, the final sample cannot be considered as fully representative for an electorate. The online link to participate in the experiment is distributed mostly by student networks. This resulted in a relatively young (average age of 23.7 years old) and highly educated (87.8% at least a bachelor degree) sample. For example, the Dutch electorate has an average age of 43.1 years old and 26.8% having at least a bachelor degree (Centraal bureau voor de statistiek, 2017). Furthermore, comparing the sample with the Dutch electorate, the sample exists of relatively many men (62.7% versus 49.3%), relatively many subjects who voted last election (94.2% versus 81.9%) and relatively many members of a political party (10.1% versus 1.7%) (Centraal bureau voor de statistiek, 2017)[17]. Also, about 11.5% of the sample is not allowed to vote in the Netherlands. These subjects are also not representative of the Dutch electorate. Furthermore, the number of candidates running in this hypothetical election makes it difficult to compare these results with the Dutch election. Dutch people are used to vote for one candidate out of multiple parties and hundreds of candidates. During the Dutch elections of March 2017, a total of 28 different parties participated in the election (Kiesraad, 2017). For this reason, an unusual election ballot during an experiment could influence the results. Future research, therefore, could try to create a sample which is more representative for the electorate of a particular country and use a familiar ballot for that same electorate.

5.2 Conclusion and Implications

The limitations of the previous section show that there is room to improve the experimental design. However, the results of this research contribute to the voting-induced polarisation literature and are very useful to build on in future research.

[17] Not all the numbers of the Dutch election of March 2017 are already available. For this reason some of the numbers are from previous Dutch elections.

First of all, the results show that voting-induced polarisation exists when people vote for one candidate. This was an important hypothesis that needed to hold before the difference in voting systems could be observed. Despite the limitations of the research, the significance of the results for this hypothesis is strong enough to conclude that a voting system in which people vote for one candidate increases the polarisation of the electorate. This result is in line with the majority of the literature which is described in Section 2.3.1. This increase in polarisation could have massive consequences in an economic, social and political way.

Although polarisation results from multiple other factors than voting itself, it is important to investigate potential solutions to decrease the level of polarisation. According to the choice induced attitude change theory, one could expect a smaller increase in polarisation when people are allowed to choose two alternatives and therefore vote less strategically or compromisingly. The results of this research indeed show a lower increase for the treatment group where people vote for two candidates. However, the results are not significant at a 5% significance level. Nevertheless, the insignificant results do not mean that the results are useless. The non-parametric tests, the parametric tests and the regressions for both different measures of spread all show a clear difference between subjects in the second treatment group and the other subjects. Furthermore, some of the results were significant at a 10% significance level, which indicates that there could be a significant relationship between different voting systems and voting-induced polarisation. Future research can build on these results by improving the limitations as mentioned in Section 5.1. The obtained results of such an improved future research could be beneficial to conclude whether voting for two candidates indeed increases polarisation within the electorate less compared to a vote-for-1 system.

In conclusion, this research investigated the relationship between voting systems and voting-induced polarisation. The results of the analyses show that, after creating cognitive dissonance, subjects change their attitude. This change is induced by voting for one candidate and lead to an increase in polarisation within the electorate. However, based on the results of the analyses it cannot be concluded at a 5% significance level that a voting system in which people vote for two candidates increases the polarisation less compared to a vote-for-1 system. Nevertheless, the results indicate that there could be a significant effect which can be obtained by improving the experimental design. Therefore, to decrease voting-induced polarisation within the electorate, more research on the effects of alternative voting systems is needed.

6. References

Aldrich, J. H. (1993). Rational choice and turnout. *American Journal of political science*, 246-278.

Anderson, C. J., Mendes, S. M., & Tverdova, Y. V. (2004). Endogenous Economic Voting: Evidence from the 1997 British Election. *Electoral Studies, 23(4),* 683-708.

Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological inquiry*, *3*(4), 303-311.

Aronson, E. (1999). Dissonance, hypocrisy, and the self-concept. *Readings about the social animal*, 219-236.'

Bakker, R., De Vries, C., Edwards, E., Hooghe, L., Jolly, S., Marks, G., ... & Vachudova, M. A. (2015). Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999–2010. *Party Politics*, *21*(1), 143-152.

Bar-Anan, Y., & Nosek, B. (2014). A Comparative Investigation of Seven Indirect Attitude Measures. *Behavior Research Methods, 46(3),* 668-688.

Beasley, R. K., & Joslyn, M. R. (2001). Cognitive Dissonance and Post-Decision Attitude Change in Six Presidential Elections. *Political Psychology, 22(3),* 521-540.

Bem, D. J. (1967). Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena. *Psychological Review, 74(3),* 183-200.

Blais, A., Vowles, J., & Aarts, K. (2002). Does the impact of polls vary across electoral systems? In *Paper, MPSA-konference*.

Bølstad, J., Dinas, E., & Riera, P. (2013). Tactical Voting and Party Preferences: A Test of Cognitive Dissonance Theory. *Political Behavior, 35(3),* 429-452.

Brehm, J. W. (1956). Postdecision Changes in the Desirability of Alternatives. *The Journal of Abnormal and Social Psychology, 52(3),* 384-389.

Centraal Bureau voor de Statistiek. (2017, February 23). *Het profiel van het electoraat in 2012 [The profile of the electorate in 2012].* From CBS: *https://www.cbs.nl/nl-nl/achtergrond/2017/08/het-profiel-van-het-electoraat-in-2012.*

Centraal Bureau voor de Statistiek. (2017, May 16). *Bevolking; Hoogstbehaald Onderwijsniveau en Onderwijsrichting [Population: Highest Attained Education and education direction].* From StatLine: http://statline.cbs.nl/statweb/publication/?vw=t&dm=slnl&pa=82816ned&d1=0&d2=a&d3=0&d4=0&d5=a&d6=0&d7=39,44,49,54,59&hd=151215-1533&hdr=g3.

Chartrand, T. L., & Bargh, J. A. (2002). Nonconscious motivations: Their activation, operation, and consequences.

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language, 35(1),* 26-58.

Coate, S., & Conlin, M. (2004). A group rule–utilitarian approach to voter turnout: theory and evidence. *The American Economic Review*, *94*(5), 1476-1504.

Crosby, L. A., & Taylor, J. R. (1983). Psychological Commitment and Its Effects on Post-Decision Evaluation and Preference Stability among Voters. *Journal of Consumer Research, 9(4),* 413-431.

Crowder, R. G. (1967). Short-term memory for words with a perceptual-motor interpolated activity. *Journal of Verbal Learning and Verbal Behavior*, *6*(5), 753-761.

Delavande, A., & Manski, C. F. (2015). Using elicited choice probabilities in hypothetical elections to study decisions to vote. *Electoral studies*, *38*, 28-37.

De Vries, C. E., Hakhverdian, A., & Lancee, B. (2013). The dynamics of voters' left/right identification: The role of economic and cultural attitudes. *Political Science Research and Methods*, *1*(02), 223-238.

Duclos, J. Y., Esteban, J., & Ray, D. (2004). Polarization: concepts, measurement, estimation. *Econometrica*, *72*(6), 1737-1772.

Egan, L. C., Santos, L. R., & Bloom, P. (2007). The origins of cognitive dissonance evidence from children and monkeys. *Psychological science*, *18*(11), 978-983.

Elinder, M. (2012). Correcting Mistakes: Cognitive Dissonance and Political Attitudes in Sweden and the United States. *Public Choice, 153(1),* 235-249.

Esteban, J.-M., & Ray, D. (1994). On the Measurement of Polarization. *Econometrica, 62(4),* 819-851.

Farrell, D. M. (2011). *Electoral systems: a comparative introduction*. Palgrave macmillan.

Festinger, L. (1957). A Theory of Cognitive Dissonance. Evanston, Illionois: Row, Peterson & Company.

Festinger, L. (1964). Conflict, decision, and dissonance (Vol. 3). *Stanford University Press.*

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*(2), 203.

Gabel, M. J., & Huber, J. D. (2000). Putting Parties in their Place: Inferring Party Left-Right Ideological Positions from Party Manifestos Data. *American Journal of Political Science*, 44(1), 94-103.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, *42*(3), 237-288.

Grossman, G. M., & Helpman, E. (2001). *Special interest politics*. MIT press.

Hariharan, P. C., & Pople, J. A. (1973). The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, *28*(3), 213-222

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, *42*(4), 1009-1055.

Heaney, M. T. (2017). Activism in an Era of Partisan Polarization. *PS: Political Science & Politics*, *50*(4).

Hogg, M. A., & Cooper, J. (2003). *The Sage handbook of social psychology*. Sage.

Huber, G., Gerber, A., & Sempolinski, J. (2009). Using ballot order to test for cognitive dissonance: *Results from a natural experiment.*

Kanbur, R. (2001). Economic Policy, Distribution and Poverty: The Nature of Disagreements. *World Development, 29(6),* 1083-1094.

Kierzenkowski, R., Pain, N., Rusticelli, E., & Zwart, S. (2016). The economic consequences of Brexit: a taxing decision. *OECD Economic Policy Papers*, (16), 1.

Kiesraad. (2017, March 21). Officiële *uitslag Tweede kamerverkiezing 15 maart 2017 [Official result Dutch National Election 15th of March 2017].*
https://www.kiesraad.nl/actueel/nieuws/2017/03/20/officiele-uitslag-tweede-kamerverkiezing-15-maart-2017

LeDuc, L., Clarke, H. D., Jenson, J., & Pammett, J. H. (1984). Partisan instability in Canada: Evidence from a new panel study. *American Political Science Review*, *78*(02), 470-484.

Lehmann, P., & Schultze, H. (2012). The MARPOR Party-Voter Data Set. *Studies [producer and distributor]*.

Lijphart, A., & Grofman, B. (1984). Choosing an electoral system. *Issues and Alternatives, New York*.

Lo, J., Proksch, S. O., & Gschwend, T. (2014). A common left-right scale for voters and parties in Europe. *Political Analysis*, *22*(2), 205-223.

McDermott, M. L. (1998). Race and gender Cues in Low-Information Elections. *Political Research Quaterly, 51(4),* 895-918.

McGregor, R. M. (2013). Cognitive dissonance and political attitudes: The case of Canada. *The Social Science Journal*, *50*(2), 168-176.

McPhee, P. (1989). The French Revolution, Peasants, and Capitalism. *The American Historical Review, 94(5)*, 1265-1280.

Morgan, J., & Stocken, P. C. (2008). Information Aggregation in Polls. The American Economic Review, 98(3), 846-896.
Mikhaylov, S., Laver, M., & Benoit, K. (2008, April). Coder reliability and misclassification in comparative manifesto project codings. In *66th MPSA Annual National Conference* (pp. 3-6).

Mullainathan, S., & Washington, E. (2009). Sticking with Your Vote: Cognitive Dissonance and Political Attitudes. *American Economic Journal: Applied Economics, 1(1),* 86-111.

Norris, P. (1997). Choosing electoral systems: proportional, majoritarian and mixed systems. *International political science review*, *18*(3), 297-312.

Norris, P. (2004). Electoral engineering: Voting rules and political behavior. *Cambridge University press*.

Pappi, F. U., & Thurner, P. W. (2002). Electoral behaviour in a two-vote system: Incentives for ticket splitting in German Bundestag elections. *European Journal of Political Research*, *41*(2), 207-232.

Powell, R., Prendergast, S., & Sharma, R. (2017). The Impact of Economic Nationalism in Europe on the Returns to Rivals of Crossborder M&A Bids.

Richardson, B. M. (1991). European Party Loyalties Revisited. *American Political Science Review, 85(3),* 751-775.

Rodríguez-Álvarez, C. (2006). Candidate Stability and Voting Correspondences. Social Choice and Welfare, 27(3), 545-570.

Rule, W. (1987). Electoral systems, contextual factors and women's opportunity for election to parliament in twenty-three democracies. *Western Political Quarterly*, *40*(3), 477-498.

Saleem, S. (2017). Turkey Referendum: New Phase in Turkish-European Relations.

Saltzman, J. (2016, June 28). The High Cost of Political Polarization. Retrieved July 20, 2016, *from Political Storm:* http://www.politicalstorm.com/high-cost-political-polarization.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological review*, *103*(2), 219.

Slapin, J. B., & Proksch, S. O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, *52*(3), 705-722.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, *72*(5), 923-955.

Voice of America. (2009). From *Voice of America:* http://www.voanews.com/a/a-13-a-2001-09-21-14-bush-66411197/549664.html

Vossler, C. A., & Kerkvliet, J. (2003). A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum. *Journal of Environmental Economics and Management*, *45*(3), 631-649.

Wilks, I. (1975). Asante in the Nineteenth Century: The Structure and Evolution of a Political Order (African Studies). *Cambridge: Cambridge University Press*.

Worchel, S., Iuzzini, J., Coutant, D., & Ivaldi, M. (2000). A multidimensional model of identity: Relating individual and group identities to intergroup behavior. *Social identity processes: Trends in theory and research*, 15-32.

Zanna, M. P., & Rempel, J. K. (1988). Attitudes: A new look at an old concept.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75-98.

Appendix A: The candidates

The Manifesto Project analyses policy statements of parties in manifestos applying a content analytical method. This content is divided into statements identifying political arguments. In total 56 policy categories are used to classify each statement (Lehman & Schultze, 2012). Budge (2013) continued with a measurement to use this statement to place candidates or parties in a political field. From the, in the meantime, 57 categories Budge (2013) placed 13 in the left category and 13 in the right category. The political position can be calculated as follows:

(R-L)/(R+L+O)

Where R represents the Right statements, L the Left statements and O the other statements. The political scale of the Manifesto Project, therefore, spans from -100 (left) to +100 (right). Nevertheless, up to 2012, the empirically observed span ranges between -75 and +65 (Lehman & Schultze, 2012). To ensure an equal dividing and using an empirically credible span the candidates are divided as shown in Table 6.

Candidate	Left statements	Right statements	Other statements	Calculation	Place
Left	3		2	(0-3)/(0+3+2)	-60%
Middle-left	1		4	(0-1)/(0+1+4)	-20%
Middle-Right		1	4	(1-0)/(1+0+4)	+20%
Right		3	2	(3-0)/(3+0+2)	+60%

Table 6: Candidate placement

Table 6 shows that the Left candidate performs three left statements and two other statements based on the Manifesto Project described by Budge (2013). These statements result in a political placement of -60% in a field which spans from -100% (left) to +100% (right). The calculated number of the other candidates can be observed by Table 6.

Furthermore, the statements of the candidates can be divided into seven policy domains (Mikhaylov et al., 2008). To ensure that candidates are comparable to one another only two comprehensive and personally interpretable domains are used in this experiment (Kanbur, 2001). Economic statements and welfare statements seem of similar importance to the potential subjects of this experiment. For this reason, every candidate makes three economic statements and two welfare and quality of life

statements. Lastly, every candidate makes at least one statement which costs extra money and one which saves extra money. Table 7 shows the different statements used in this experiment.

	Statement	Right/left/other	Economy/Welfare	Candidate
1	More market regulation	Left	Economy	Left and Middle-Left
2	More controlled economy	Left	Economy	Left
3	Welfare state expansion	Left	Welfare	Left
4	More financial incentives	Right	Economy	Right and Middle-right
5	Free market economy	Right	Economy	Right
6	Welfare state limitation	Right	Welfare	Right
7	Less to technology and infrastructure	Other	Economy	Left and Middle-left
8	More Corporatism	Other	Economy	Middle-left
9	Free enterprise	Other	Economy	Middle-right
10	More to technology and infrastructure	Other	Economy	Right and Middle-right
11	More to culture	Other	Welfare	Left and Middle-left
12	More Social justice	Other	Welfare	Middle-left
13	Environmental protection	Other	Welfare	Middle-right
14	Less to culture	Other	Welfare	Right and Middle-right

Table 7: Statements of the candidates

The first three statements of table 7 are left categorised statements, the next three statements are right categorised statements, and the last eight statements are not left and not right categorised. The statements which are categorised as 'other' are chosen because I expect subjects to interpret these statements somewhat left and somewhat right. The control question about the political placement is needed to check whether subjects place the candidates as for where this experiment the candidates aims.

Appendix B: Results alternative measure of spread

This appendix gives an overview of the nonparametric and parametric test results of the alternative measure of spread. Table 8 shows the results of the one sample tests and Table 9 lists the results of the between sample tests. The two sample t-tests are conducted with an unequal variance t-test.

Test	Score	95% confidence interval/ variance	P-value
T-test Control group	T= -0.325	-3.593 to 2.593	0.746
T-test Treatment group 1	T= -0.370	-3.427 to 2.371	0.714
T-test Treatment group 2	T= 2.116	0.156 to 5.807	0.039
Wilcoxon signed-rank test Control group	W= 161.5	1218.750	0.753
Wilcoxon signed-rank test Treatment group 1	W= 284.5	4032.875	0.450
Wilcoxon signed-rank test Treatment group 2	W= 685	7822.875	0.059

Table 8: Statistical results one sample tests using the alternative measure of spread

Test	U-score/ difference	Z-score	p-value
Kruskal-Wallis test	3.555		0.169
Mann-Whitney U test CG – TG 1	941.500	0.701	0.481
Mann-Whitney U test CG – TG 2	862.000	-1.039	0.297
Mann-Whitney U test TG 1 – TG 2	1043.000	-1.831	0.066
Two-sample T-test CG – TG 1	-0.028	-0.013	0.989
Two-sample T-test CG – TG 2	-3.510	-1.749	0.084
Two-sample T-test TG 1 – TG 2	4.249	-1.669	0.098

Table 9: Statistics results between sample tests alternative measure of spread